

1 D. John Sauer, Mo. Bar No. 58721*
2 Justin D. Smith, Mo. Bar No. 63253*
3 James Otis Law Group, LLC
4 13321 North Outer Forty Road, Suite 300
5 St. Louis, Missouri 63017
6 Telephone: (314) 562-0031
7 John.Sauer@james-otis.com

8 *Attorneys for Intervenor-Defendants President Petersen and Speaker Toma*

9 **IN THE UNITED STATES DISTRICT COURT**
10 **FOR THE DISTRICT OF ARIZONA**
11 **TUCSON DIVISION**

12 Jane Doe, *et al.*,

13 Plaintiffs,

14 v.

15 Thomas C. Horne, in his official capacity
16 as State Superintendent of Public
17 Instruction, *et al.*,

18 Defendants.
19
20
21

Case No. 4:23-cv-00185-JGZ

**Intervenor-Defendants' Expert
Declarations and Exhibits for
Preliminary Injunction Hearing**

22
23 Intervenor-Defendants President Petersen and Speaker Toma, pursuant to this
24 Court's June 14, 2023 Minute Entry, respectfully submit this statement identifying
25 Intervenor-Defendants' expert declarations and documentary exhibits they intend to rely
26 upon for the upcoming Preliminary Injunction hearing. Filed with this statement are the
27 following:

- 28 1. Declaration of Dr. Greg Brown (previously filed as Doc. 82-1)

2. Declaration of Dr. James Cantor (previously filed as Doc. 82-2)
3. Declaration of Dr. Chad Carlson (previously filed as Doc. 82-3)
4. Supplemental Declaration of Dr. Greg Brown
5. Supplemental Declaration of Dr. James Cantor
6. Supplemental Declaration of Dr. Chad Carlson
7. Catley, M. and G. Tomkinson, Normative health-related fitness values for children: analysis of 85,437 test results on 9-17-year-old Australians since 1985. Br. J. Sports Med. published online October 21, 2011. Bjsm.bmj.com
8. Davis, S.M., et al., Sex Differences in Infant Body Composition Emerge in First 5 Months of Life. J. Pediatr. Endocrinol. Metab. 32(11): 1235–1239 (2019)
9. DiMiguel-Etayo P., Physical Fitness Reference Standards in European Children: The IDEFICS Study. Int J Obes (Lond) 38(2):557-566 (2014)
10. Dohrmann P., Throwing and Kicking Ability of 8 Year Old Boys and Girls. The Research Quarterly 35(4):464-471 (1964)
11. Eiberg, S. et al, Maximum oxygen uptake and objectively measured physical activity in Danish children 6-7 years of age: the Copenhagen school child intervention study. Br J Sports Med 39:725-30 (2005)
12. Gulias-González, et al. (2014). Physical fitness in Spanish schoolchildren aged 6–12 years: Reference values of the battery EUROFIT and associated cardiovascular risk. Journal of School Health, 84, 625–635
13. Kasovic M. et al., Secular Trends in Health-Related Physical Fitness Among 11-14 Year-Old Croatian Children and Adolescents From 1999 to 2014. Scientific Rep 11:11039 (2021)
14. Kirchengast, S., Sexual dimorphism in body composition, weight status and growth in prepubertal school children from rural areas of eastern Austria. Collegium Antropologicum 25(1):21-30 (2001)
15. Lombardo MP et al. On the evolution of the sex differences in throwing:

- 1 Throwing is a male adaptation in humans. *The Quarterly Review of Biology*
2 93(2):91-119 (2018)
- 3 16. Moreland E, Cheung AS, Hiam D, Nolan BJ, Landen S, Jacques M, Eynon
4 N, Jones P. Implications of gender-affirming endocrine care for sports
5 participation. *Ther Adv Endocrinol Metab.* 2023 Jun
6 8;14:20420188231178373.
7 [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10262668/pdf/10.1177_20](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10262668/pdf/10.1177_204201882311783.pdf)
8 204201882311783.pdf
- 9 17. Sauka, M., et al. (2011). Physical fitness in northern European youth:
10 Reference values from the Latvian Physical Health in Youth Study.
11 *Scandinavian Journal of Public Health*, 39, 35–43
- 12 18. Tambalis K et al., Physical Fitness Normative Values for 6-18 Year Old
13 Greek Boys and Girls. *Eur J Sport Sci* 16(6):736-46 (2016)
- 14 19. Taylor, R.W., et al., Gender differences in body fat content are present well
15 before puberty. *Int. J. Obes. Relat. Metab. Disord.* 21(11): 1082-4 (1997)
- 16 20. Taylor, R.W., et al., Sex differences in regional body fat distribution from
17 pre- to postpuberty. *Obesity* 18(7): 1410-16 (2010)
- 18 21. Thomas, J.R. and K. E. French, Gender differences across age in motor
19 performance: a meta-analysis. *Psych. Bull.* 98(2):260-282 (1985)
- 20 22. Tønnessen, E. et al., Performance development in adolescent track & field
21 athletes according to age, sex, and sport discipline. *PLoS ONE* 10(6):
22 e0129014 (2015)
- 23 23. UK Sports Councils' Equality Group, International Research Literature
24 Review,
25 [https://equalityinsport.org/docs/300921/Transgender%20International%20](https://equalityinsport.org/docs/300921/Transgender%20International%20Research%20Literature%20Review%202021.pdf)
26 [Research%20Literature%20Review%202021.pdf](https://equalityinsport.org/docs/300921/Transgender%20International%20Research%20Literature%20Review%202021.pdf) (2021)
- 27 24. Woll, A., et al. (2011). The 'Motorik- Modul' (MoMo): Physical fitness and
28 physical activity in German children and adolescents. *European Journal of*

- 1 Pediatrics, 170, 1129–1142
- 2 25. Zheng Y, Ye W, Korivi M, Liu Y, and Hong F. Gender Differences in
- 3 Fundamental Motor Skills Proficiency in Children Aged 3-6 Years: A
- 4 Systematic Review and Meta-Analysis. *Int J Environ Res Public Health* 19,
- 5 2022
- 6 26. American Psychiatric Association (2022). Diagnostic and statistical manual
- 7 for mental disorders, fifth edition, text revision (DSM-5-TR). Washington,
- 8 DC: Author. (page 511 excerpt)
- 9 27. Ellis Barrera et al., The Medical Implications of Banning Transgender Youth
- 10 from Sport Participation, 176(3) *JAMA Pediatrics* 223-24 (2022)
- 11 28. Handelsman, D. et al., Circulating testosterone as the hormonal basis of sex
- 12 differences in athletic performance. *Endocrine Reviews* 39(5):803-829 (Oct
- 13 2018)
- 14 29. Handelsman, D. Sex differences in athletic performance emerge coinciding
- 15 with the onset of male puberty, 87 *Clinical Endocrinology* 68 (2017)
- 16 30. Hughes, L. et al., Pediatric Provider Perspectives on Laws and Policies
- 17 Impacting Sports Participation for Transgender Youth, 9(4) *LGBT Health*
- 18 247-53 (2022)
- 19 31. McManus, A. and N. Armstrong, Physiology of elite young female athletes.
- 20 *J Med & Sport Sci* 56:23-46 (2011)
- 21 32. Klaver M., Early hormonal treatment affects body composition and body
- 22 shape in young transgender adolescents. *J Sex Med.* 15(2):251-260 (2018)
- 23 33. Tack, L.J.W. et al., Proandrogenic and antiandrogenic progestins in
- 24 transgender youth: differential effects on body composition and bone
- 25 metabolism. *J. Clin. Endocrinol. Metab*, 103(6):2147-56 (2018)
- 26 34. Riley Gaines Testimony before the Senate Judiciary Committee, June 21,
- 27 2023. Independent Women’s Voice.
- 28

1 Dated: June 29, 2023

Respectfully submitted,

3 JAMES OTIS LAW GROUP, LLC

4 /s/ Justin D. Smith

5 D. John Sauer, Mo. Bar No. 58721*

6 Justin D. Smith, Mo. Bar No. 63253*

13321 North Outer Forty Road, Suite 300

7 St. Louis, Missouri 63017

8 (816) 678-2103

Justin.Smith@james-otis.com

9 * *pro hac vice*

10 *Attorneys for Proposed Intervenor-Defendants*

CERTIFICATE OF SERVICE

I hereby certify that, on June 29, 2023, I caused a true and correct copy of the foregoing to be filed by the Court's electronic filing system, to be served by operation of the Court's electronic filing system on counsel for all parties who have entered in the case.

/s/ Justin D. Smith

1 D. John Sauer, Mo. Bar No. 58721*
2 Justin D. Smith, Mo. Bar No. 63253*
3 James Otis Law Group, LLC
4 13321 North Outer Forty Road, Suite 300
5 St. Louis, Missouri 63017
6 Telephone: (314) 562-0031
7 John.Sauer@james-otis.com

8 *Attorneys for Intervenor-Defendants President Petersen and Speaker Toma*

9 **IN THE UNITED STATES DISTRICT COURT**
10 **FOR THE DISTRICT OF ARIZONA**
11 **TUCSON DIVISION**

12 Jane Doe, *et al.*,

13 Plaintiffs,

14 v.

15
16
17
18
19 Thomas C. Horne, in his official capacity
20 as State Superintendent of Public
21 Instruction, *et al.*,

22 Defendants.
23

Case No. 4:23-cv-00185-JGZ

**Rebuttal Declaration of Dr. Gregory A.
Brown, Ph.D., FACSM, in Further
Support of Intervenor-Defendants'
Opposition to Plaintiffs' Motion for a
Preliminary Injunction**

24 I, Gregory A. Brown, declare as follows:

- 25 1. I have submitted an initial declaration to this Court dated May 18, 2023.
26 2. I now submit this expert rebuttal declaration based on my personal
27 knowledge, and it reflects my expert opinions.
28 3. In preparing this rebuttal declaration, I have reviewed the expert declarations

DEFENDANT'S
EXHIBIT

4

1 filed by Plaintiffs, submitted by Dr. Shumer and Dr. Budge.

2 4. In Dr. Shumer's rebuttal to my expert declaration, at paragraph 4, he states
3 "the studies and findings discussed throughout Dr. Brown's declaration support the
4 scientific consensus that the biological cause of average group differences in athletic
5 performance between males and females is the rise in circulating levels of testosterone
6 beginning in endogenous male puberty." This statement seems to completely ignore
7 paragraphs 77-115 of my declaration and the data tables contained therein along with the
8 data tables included in the appendix (pages 99-107), all of which are drawn from 16
9 separate sources, which document numerous differences in physical fitness and athletic
10 performance between boys and girls before the onset of puberty.

11 5. In Dr. Shumer's rebuttal to my expert declaration, at paragraph 5, he states
12 that I have misrepresented the writing of McManus and Armstrong (2011) when I wrote
13 (in paragraph 77) "It is often said or assumed that boys enjoy no significant athletic
14 advantage over girls before puberty. However, this is not true. Writing in their seminal
15 work on the physiology of elite young female athletes, McManus and Armstrong (2011)
16 reviewed the differences between boys and girls regarding bone density, body composition,
17 cardiovascular function, metabolic function, and other physiologic factors that can
18 influence athletic performance. They stated, 'At birth, boys tend to have a greater lean
19 mass than girls. This difference remains small but detectable throughout childhood with
20 about a 10% greater lean mass in boys than girls prior to puberty.' (28) 'Sexual dimorphism
21 underlies much of the physiologic response to exercise,' and most importantly these
22 authors concluded that, 'Young girl athletes are not simply smaller, less muscular boys.'
23 (23)." Dr. Shumer faults me for not noting that the McManus paper found no difference
24 between the sexes in measures of *some other physical characteristics*. But I never claimed
25 that prepubertal boys and girls are physically different in *every* respect. What I claimed—
26 and what the McManus citation supports—is that prepubertal boys and girls are different
27 in *some* areas that contribute to athletic performance. McManus found measurable
28 differences between prepubertal boys and girls in body fat mass, percent body fat, lean

1 body mass, peak oxygen uptake, maximal pulmonary ventilation, blood volume, cardiac
2 function.

3 6. I would therefore like to provide the following further quotations from
4 McManus and Armstrong supporting my reading of the paper that boys enjoy a significant
5 athletic advantage over girls before puberty.

6 7. “Small sex differences in fat mass and percent body fat are evident from mid-
7 childhood...” (at 27) – Mid childhood is considered to be ages 6-12. This statement is used
8 by Dr. Shumer in an endeavor to discredit my expert report, when indeed it supports my
9 report. “Small differences” is an ambiguous term, yet athletic advantages are often the sum
10 of many small differences (as pointed out in my declaration, differences of 3-5% are often
11 more than the difference between a gold medal and no medal, see paragraphs 111-112).
12 Furthermore, the magnitude of an advantage is not a deciding factor in whether that
13 advantage is or is not allowed in sports. Anabolic-Androgenic steroids provide a 5-20%
14 advantage in muscle *strength* (Hartgens and Kuipers, 2004) and are almost universally
15 banned as *performance* enhancing substances. Androstenedione was sold as a testosterone
16 enhancing nutritional supplement in the late 1990s and early 2000s and was banned as a
17 performance enhancing substance even though research shows that androstenedione intake
18 does not enhance the adaptations to resistance training (King et al. 1997, Brown et al.
19 2000). Fastskin swimming suits provide a $3.2 \pm 2.4\%$ performance benefit in swimming
20 (Chatard and Wilson, 2008), and are banned from use by FINA.

21 8. “At birth, boys tend to have a greater lean mass than girls. This difference
22 remains small but detectable throughout childhood with about a 10% greater lean mass in
23 boys than girls prior to puberty.” (at 28)

24 9. “In comparison to boys, girls are characterised with a smaller absolute peak
25 VO_2 . Predicted values range from 1.5 to 2.2 litres $\cdot\text{min}^{-1}$ in 10- to 16-year-old girls and are
26 lower than boys by 11, 19, 23 and 27% at ages 10, 12, 14 and 16 years of age, respectively.”
27 (at 30) Peak VO_2 is an estimation of maximal oxygen consumption (called $\text{VO}_{2\text{max}}$),
28 which accounts for 30-40% of performance in endurance exercise. Puberty is not typically

1 experienced by boys or girls by 10 years of age.

2 10. “In children, like adults, exercise pulmonary gas exchange depends on
3 pulmonary ventilation (VE) and at maximal work rates high rates of ventilation are usual.
4 Maximal values of 49– 95 litres•min⁻¹ have been recorded for girls between the ages of 9
5 and 16 years [] and there is a consistent sex difference with values somewhat higher in
6 boys (58– 105 litres • min⁻¹) for the same age span.” “Maximum ventilation remains
7 higher in boys, whether controlled for body size using a ratio standard or allometric
8 adjustment with either stature and/or body mass []. Thus, the higher peak VO₂ in boys is
9 indeed supported by a higher VE.” (at 31)

10 11. When describing differences in blood volume per unit of body mass: “When
11 normalised using a ratio standard with body mass, differences between girls and boys were
12 apparent from about 6 years of age, with values lower in the girls.” (at 32)

13 12. “There are clear differences in cardiac function at rest and during exercise
14 between girls and boys, with differences apparent even prior to puberty. The electrical
15 conduction system is influenced by sex steroid hormones, with girls normally having
16 higher resting heart rates than boys – somewhere in the magnitude of 90 beats per minute
17 at around 10-12 years of age []. This is thought to relate to intrinsic differences in the sinus
18 node pacemaker [], a difference notable at birth with newborn boys displaying lower
19 baseline heart rates than girls []. The higher resting heart rate in girls is often explained as
20 an artefact of differences in cardiac dimensions, and indeed the ratio of heart mass to body
21 mass has been found to be higher in boys than girls at birth, remaining so through
22 adolescence []. Heart volume has also been found to be greater in boys with values of 342
23 and 403 ml for pre-pubertal girls and boys, respectively...” (at 32)

24 13. “Data recently published from a thoracic impedance measure of peak
25 C[ardiac]I[ndex] and MRI markers of cardiac size [] demonstrated that pre- pubertal boys
26 had a 16.7% higher (a- v O₂) difference than girls.” (at 34) – Cardiac index is an assessment
27 of the cardiac output value based on the patient’s size. Cardiac output is the volume of
28 blood the heart pumps per minute. (a-v O₂) difference is the arterio-venous oxygen

1 difference, and measures how well the tissues extract oxygen from the blood stream. (a-v
2 O₂) difference accounts for roughly 40-50% of maximal oxygen consumption.

3 14. “Results showed phase II pVO₂ kinetics were approximately 20% slower in
4 pre- pubertal girls compared to boys ... This is suggestive of a lower tolerance of fatigue
5 in the girls” (at 35) – pVO₂ stand for Pulmonary Oxygen Uptake, and pVO₂ kinetics
6 provides an insight into the integrated capacity of an organism to transport and utilize
7 oxygen to support an increased rate of energy turnover in contracting muscle cells.

8 15. “To summarise, there are differences between boys and girls in the aerobic
9 responses to exercise which cannot be accounted for solely by size.” (at 35)

10 16. Dr. Shumer states (at paragraph 5) that the article by McManus and
11 Armstrong is published in the journal *Medicine and Science in Sports and Exercise* (which
12 is the flagship journal for the American College of Sports Medicine). The referenced article
13 by McManus & Armstrong is actually published in *Medicine and Sport Science*, which is
14 a book series (not a journal) and is not in any way affiliated with the American College of
15 Sports Medicine.

16 17. At paragraph 6, Dr. Shumer states “Dr. Brown gives the false impression that
17 all 22 of the peer-reviewed publications demonstrated differences on total body fat. Instead,
18 Staiano and Katzmarzyk expressly note that ‘not all studies demonstrate sex differences in
19 T[otal]B[ody]F[at] before puberty.’” Dr. Shumer contends that my report is deceptive
20 because Staiano’s conclusion—that prepubertal girls tend to have more body fat (which is
21 exactly what the article says: “In prepubertal children, girls typically have more T[otal]
22 B[ody] F[at] than boys.”)—was not based on unanimous evidence, but rather on the weight
23 of the evidence. Staiano noted that, of the 22 studies reviewed, four of them found similar
24 body fat between boys and girls. Staiano suggested that these studies were influenced by a
25 failure to control for “other influences like age, maturational status and obesity status.” In
26 any event, I did not claim that the evidence was unanimous; I simply cited the peer-
27 reviewed conclusion reached by Staiano based on 18 of the 22 studies Staiano reviewed.
28 That isn’t deceptive. And experts do not need unanimity to reach a reliable conclusion;

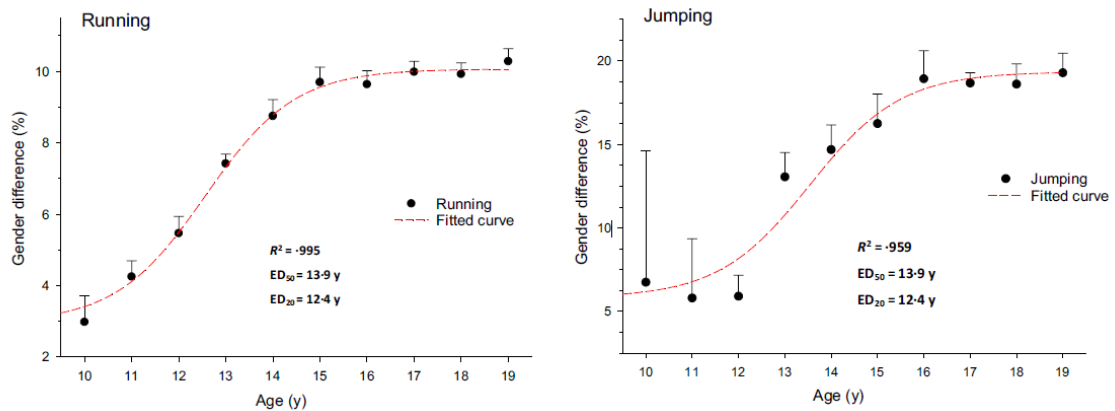
1 rather, they are to look to the great weight of the evidence, which is exactly what I did.

2 18. In paragraphs 7 and 8, Dr. Shumer criticizes my partial use of a statement
3 from Handelsman: “Dr. Brown further misrepresents Handelsman (2018)’s findings,
4 notably omitting key portions from the study he cites. Dr. Brown writes, ‘[t]here is
5 convincing evidence that the sex differences in muscle mass and strength are sufficient to
6 account for the increased strength and aerobic performance of men compared with women
7 and is in keeping with the differences in world records between the sexes.’ (Brown Decl.
8 ¶ 59; Brown Hecox Decl. ¶ 88.) But Dr. Brown omits the following sentence from
9 Handelsman which explains that ‘[t]he basis for the sex difference in muscle mass and
10 strength is the sex difference in circulating testosterone.’ David Handelsman, et al.
11 Circulating Testosterone as the Hormonal Basis of Sex Differences in Athletic
12 Performance, 39 Endocrine Revs. 803, 816 (2018) (emphasis added).” The second half of
13 that sentence is purposefully omitted as I do not agree with the proposition that testosterone
14 is the only factor responsible for sex-based differences in athletic performance. Indeed, I
15 dedicate many pages of my expert report to demonstrating that there are sex-based
16 differences in athletic performance before puberty citing numerous sources and providing
17 many tables of data. In many places within my report, I acknowledge that puberty driven
18 increases in testosterone in males causes large increases in the differences in athletic
19 performance between males and females (for example, see paragraphs 126-130), so to omit
20 a partial sentence from a single source is hardly misleading.

21 19. I would like to point out that after paragraph 127, I include the lower panel
22 of figure 2 from Handelsman (2017) which shows “Fitted sigmoidal curve plot of gender
23 differences in performance (in percentage) according to age (in years) in running, jumping
24 and swimming events as well as serum testosterone. Data shown as mean and standard
25 error of the mean of the pooled gender differences by age.”

26 20. I would like to add the upper panel to figure 2 from Handelsman (2017) (see
27 below), which shows Gender differences in performance (in percentage) according to age
28 (in years) in running events including 50 m, 60 m, 100 m, 200 m, 300 m, 400 m, 500 m,

600 m, 800 m, 1000 m, 1500 m, 1 mile, 2000 m, 3000 m and 2 miles (upper left panel) and in jumping events including high jump, pole vault, triple jump, long jump and standing long jump (upper right panel). This figure demonstrates an average male performance advantage of ~3% in running at age 10, ~4% at age 11, and ~5% at age 12, and this figure also demonstrates an ~6% male advantage in jumping at age 10, and ~5% at ages 11 and 12. Extrapolating the error bars in these graphs (which represent the standard deviation, it is very reasonable to expect that a majority of boys at ages 10, 11, and 12 will outperform girls of the same age.



21. In paragraphs 9-12, Dr. Shumer refers to an article written for “The Conversation” (a network of not-for-profit media outlets publishing news stories and research reports online, with accompanying expert opinion and analysis written by academics and researchers) summarizing “research published in American Academy of Neurology Journal.” It is important to note that the research published in the journal *Neurology*, published by the American Academy of Neurology, referenced by Dr. Shumer evaluated “Twelve functional outcome measures were collected from 1,000 healthy individuals aged 3-101 years”, and did not specifically focus on children or adolescents. In The Conversation, it is explained that “As part of wider research to assess people’s physical capabilities across the lifespan, we tested 300 children and adolescents between the ages of 3 and 19.”¹

¹ <https://theconversation.com/when-it-comes-to-sport-boys-play-like-a-girl-80328>

1 22. In contrast to the above citation by Dr. Shumer to a single study with “300
2 children and adolescents between the ages of 3 and 19,” my report cited peer-reviewed
3 research publications of a range of sample sizes focusing on children, not children as “part
4 of wider research to assess people’s physical capabilities across the lifespan.”

5 23. Here are some of the studies I cited:

6 a. In a test of reaction time in Spanish Preschool children (1,845 girls
7 and 1,896 boys with a mean age of 55.93 ± 11.14 months) boys performed better
8 than girls (paragraphs 43 and 114 referencing Latorre-Roman et al. 2018)

9 b. A summary of data from “a national sample of Canadian children and
10 youth” ages 7-17 years demonstrating that boys have higher aerobic power than girls
11 of the same age (Paragraph 80, Citing Malina et al. 2004, and Gauthier et al. 1983).

12 c. In an evaluation “of 703 male and female elite young German athletes
13 aged 8-18” (1) “fitness development precedes sports specialization” (2) and “males
14 outperformed females in C[ounter]M[ovement]J[ump], D[rop]J[ump],
15 C[hange]o[f]D[irection speed] performances and hand grip strength.” (Paragraph 81
16 Citing Lesinski et al. 2020).

17 d. A total of 424,328 Greek boys and girls aged 6-18 years using
18 standing long jump to measure lower body explosive power, sit and reach to
19 measure flexibility, timed 30 second sit ups to measure abdominal and hip flexor
20 muscle endurance, 10 x 5 meter shuttle run to evaluate speed and agility, and multi-
21 stage 20 meter shuttle run test to estimate aerobic performance. For each of the
22 fitness tests, performance was better in boys compared with girls, except for the sit
23 and reach test. (Paragraphs 82 -84, 100, Citing Tambalis 2016).

24 e. USA Presidential Fitness Testing data for 85th and 50th percentile
25 demonstrating that boys perform better on tests of muscle strength and running
26 endurance (Paragraph 85-86).

27 f. An evaluation of 85,000 Australian children aged 9-17 years old
28 showed that, compared with 9-year-old females, 9-year-old males were faster over

1 short sprints (9.8%) and 1 mile (16.6%), could jump 9.5% further from a standing
2 start (a test of explosive power), could complete 33% more push-ups in 30 [seconds]
3 and had 13.8% stronger grip. (Paragraphs 88-89, citing Catley & Tomkinson, 2013).

4 g. Evaluation of the “Eurofit” test battery on children from 30 European
5 countries and 2,779,165 test performances in 9-17 year old boys and girls showed
6 that boys performed better than similarly aged girls at each age on tests of muscular
7 strength, muscular endurance, and aerobic fitness (Paragraphs 90-93, citing
8 Tomkinson 2018).

9 h. An evaluation of 20m shuttle run performance in 1,142,026 children
10 aged 9-17 in 50 countries showing that boys performed better than girls of the same
11 age (paragraphs 94-95, citing Tomkinson et al., 2017).

12 i. An evaluation of 10,302 children aged 6-10.9 years of age, from the
13 European countries of Sweden, Germany, Hungary, Italy, Cyprus, Spain, Belgium,
14 and Estonia demonstrating that boys performed better than girls in speed, lower-
15 and upper-limb strength and cardiorespiratory fitness. (Paragraphs 97-99, citing De
16 Miguel-Etayo et al. 2014).

17 j. An evaluation of 18 studies for males (N=5676 in total) and 17 studies
18 for females (N=5489 in total) in the United States and Canada demonstrating the
19 boys had strength advantages of between 13 and 28 percent, with the remaining
20 outlier recording only a 4% advantage for 7-year-old boys (Paragraph 101, citing
21 Silverman 2011).

22 k. An analysis of vertical jump measurements of 7,614 healthy
23 Colombian schoolchildren aged 9 -17.9 years of age, showing that boys jump higher
24 than girls of the same age (Paragraph 103, citing Ramírez-Vélez et al, 2017).

25 l. An analysis of vertical jump measurements of 1,845 children aged 10-
26 15 years in primary and secondary schools in the East of England, showing that boys
27 jump higher than girls of the same age (Paragraph 104, citing Taylor 2010).

28 m. Data from USA Track & Field (Paragraphs 107, 108).

1 n. Data from Athletic.net for the USA (Paragraph 109, 110).

2 o. Data from 366 Danish boys and 332 Danish girls between the ages of
3 6 and 7 years old showing that boys have higher measurements of aerobic fitness,
4 even if the boys and girls engage in the same amount of physical activity (Paragraph
5 113, citing Eiberg 2005).

6 24. In Paragraph 10, Dr. Shumer contends that age, location, or socioeconomic
7 factors have not been controlled for in the above-referenced studies. This is quite simply
8 not so, as the vast majority of these papers compared the performance of children of the
9 same age (as demonstrated in the normative data presented in paragraphs 77-115 and the
10 appendix of my declaration), and, as explained above, the male advantages have been
11 documented in a wide range of countries.

12 25. To further demonstrate that prepubertal boys exhibit advantages in measure
13 of physical fitness and motor control which give them advantages in sports compared to
14 girls of the same age, here are even more papers:

15 a. A Systematic Review and Meta-Analysis of 38 articles studies were
16 carried out in 19 different countries (Australia, Belgium, Brazil, Britain, China,
17 Croatia, Germany, Iran, Indonesia, Ireland, Japan, Korea, Myanmar, Poland,
18 Portugal, Puerto Rico, Singapore, South Africa, and the USA representing data for
19 8394 children ages 3-6 years old who were assessed for object control skills.
20 Significant differences were found, favoring boys vs. girls at ages 3, 4, 5, and 6 with
21 at least some of the differences attributable to biology (Zheng et al, 2022).

22 b. 1,682 children and adolescent aged 6-17 years from central Spain,
23 divided into prepubertal and pubertal groups based on Tanner stages demonstrating
24 that pre-pubertal boys had more muscle mass, less fat mass, and performed better
25 girls on tests of countermovement jump, handgrip strength, and 20 m shuttle run
26 (Manzano-Carrasco et al. 2022).

27 c. 3,179 preschool children (1678 boys) ages 2.8-6.4 years from 10
28 different cities and towns in Spain and found boys outperformed girls in the 20 m

1 shuttle run, handgrip strength, standing long jump, and 4X 10 m shuttle run
2 (Cadenas-Sanchez, 2019).

3 d. 31,484 children (16,023 boys and 15,461 girls) ages 6-11 years old
4 from a representative sample of the French population with boys performing better
5 on tests of Cardiorespiratory fitness, muscular endurance, and speed (Vanhelst et al.
6 2020).

7 e. 341 young Nigerian children (ages 3 to 5) At each age level the boys
8 consistently performed better than the girls tests of catching, standing long jump,
9 tennis ball throw and speed run (Toriola and Ingokwe, 1986).

10 f. 434 low-income preschool children from Santiago Chile (246 boys;
11 5.48 ± 0.31 years) showing that boys were heavier and taller than girls, with boys
12 performing better on handgrip strength test, standing long jump. and 20 m sprint
13 (Cadenas-Sanchez, 2015).

14 26. It is also important to note that sports do not take into account socioeconomic
15 factors or location. For example, at a youth wrestling tournament the athletes may be
16 categorized based on sex, age, and body weight, but not socioeconomic status or location.
17 In a youth soccer tournament, the athletes may be categorized based on sex, age, or possibly
18 the team skill rating, but not socioeconomic status or location.

19 27. In Paragraph 12, Dr. Shumer claims that there has been wide replication of
20 the lack of difference in sporting performance between prepubertal boys and girls and states
21 that there is a general consensus that there are no sex-based differences in athletic
22 performance before puberty, and yet cites only two sources. Neither of these sources
23 professes to present a scientific consensus statement on the presence or lack of sex-based
24 difference in performance before puberty.

25 28. In reading Senefeld et al., Sex Differences in Youth Elite Swimming, 14
26 PLOS ONE 1, 1–2 (2019), these authors cite only two sources regarding the sex-based
27 differences in sporting performance in 10 to 12-year-olds, one of which is the Handelsman
28 (2018) paper and the other is a paper by Tonnessen et al. (Performance development in

1 adolescent track and field athletes according to age, sex and sport discipline. PloS one.
2 2015;10(6):e0129014). Indeed, Senefeld et al. state “However, the sex-based differences
3 in performance prior to age 10 are unknown ...” (at page 2) and “However it is clear that
4 these data provide one of the only examples of faster (or at least not slower) sports
5 performance for girls than boys.” (at page 8)

6 29. In paragraph 13, Dr. Shumer describes these data as “Demographic Data”,
7 which is incorrect. Demographic data are used to help understand the statistical
8 characteristics of human populations. Demographic data can contain specific information
9 about the characteristics of a given population, such as the following: age range, race and
10 ethnicity, sex, gender, level of education, income, employment status, occupation,
11 homeownership, birth rates, death rates, marriage rates, religious affiliation, political
12 affiliation, spoken language, geographic location, or hobbies and interests. Many of the
13 studies which Dr. Shumer calls demographic data are normative data, which are
14 information from a population of interest that establishes a baseline distribution of results
15 for that particular population (Lee & Schuele, Abdi & Williams, Encyclopedia of Research
16 Design, Sage Publications, 2010).

17 30. The competition data presented in my report represent the under-8 and 9 to
18 10-year-old records from USA Track & Field, and annual performance data gleaned from
19 Athletic.net for the State of Arizona in 2022, and for the entire United States in 2021.

20 31. I recently (June 2, 2023) presented research at the 2023 annual meeting of
21 the American College of Sports Medicine using nationwide results from Athletic.net
22 demonstrating that over the years 2017-2021, the top 10 boys ages 7-8 and 9-10 ran faster
23 than girls of the same ages and jumped higher and faster in 100m, 200m, 400m, 800m,
24 600m, high jump and long jump by 3-10% than the girls in every event every year (Brown
25 GA, Brown CJ, Shaw I, Shaw B. Boy and Girls Differ in Track and Field Event
26 Performance Before Puberty. 70th Annual Meeting of the American College of Sports
27 Medicine. Denver CO. Presentation 2577. May 30 – June 2, 2023). There was another
28 presentation in the same session in which the authors used data from Athletic.net for the

1 top 10 male and female athletes for the years 2019, 2020, and 2021 for ages 7-18 years,
2 and observed that prepubertal males outperformed females of the same age by 3-10% in
3 the 100m, 200m, 400m, 800, high jump and long jump every year and overall (Atkinson
4 MA, Linde JJ, Hunter SK. Sex Differences in Performance of Elite Youth Track and Field
5 Athletes. 70th Annual Meeting of the American College of Sports Medicine. Denver CO.
6 Presentation 2572. May 30 – June 2, 2023). This demonstrates that (1) data from
7 Athletic.net are considered sufficiently reliable for scholarly endeavors, and (2) prepubertal
8 male advantages in running and jumping are consistently demonstrated in elite youth.

9 32. Additionally, The Motivational Times, from USA Swimming, show under
10 10-year-old boys consistently swimming faster than under 10-year-old girls.²

11 33. In paragraphs 14-27, Dr. Shumer contends that “Transgender girls who
12 receive puberty suppressing medication at the onset of puberty have no athletic advantage
13 over other girls.” While Dr. Shumer is correct to state that there is no research showing
14 puberty suppression and or cross sex hormones does not eliminate male athletic
15 advantages, similarly there is no research showing it does. Dr. Shumer does not cite any
16 studies showing that puberty suppression results in transgender girls exhibiting athletic
17 performance that is the same as equally aged, gifted, and trained females. Dr. Shumer
18 attempts to deflect the research showing the males who take puberty blockers and/or cross
19 sex hormones retain male pattern advantages in lean body mass, muscle strength, body
20 height, and so forth, by stating that this research does not demonstrate athletic advantages.
21 In this he ignores the commonly held tenet in the professions of exercise physiology and
22 strength & conditioning that lean body mass is one the major factors driving athletic
23 performance overall, and driving the sex-based differences in athletic performance (see my
24 declaration, paragraphs 61 and 81 for explanation of this tenet).

25 34. Overall, athletes spend an inordinate amount of time in the weight room, on
26 the track, in the pool, etc. trying to improve their physical fitness, because improved

27 ² [https://swimswam.com/usa-swimming-releases-age-group-motivational-times-for-2021-](https://swimswam.com/usa-swimming-releases-age-group-motivational-times-for-2021-2024/)
28 2024/

1 physical fitness translates to improved athletic performance. Whether it is measured as a
2 higher VO₂max in an endurance athlete, a higher 1-repetition maximum for a thrower or
3 wrestler, or a larger amount of lean body mass in almost any athlete, these measures of
4 improved physical fitness are indicators of a greater potential for successful athletic
5 performance. The differences in physical fitness between males and females before and
6 after puberty predispose males to a winning performance if they were to compete against
7 females of the same age who have the same training and sports background.

8 35. Lean body mass is a significant determinant of muscle strength and sports
9 performance. As demonstrated by Almiray-Stot et al. (2022), in healthy children ages 5 to
10 19-years-old, lean body mass is significantly correlated to muscle strength in both boys
11 and girls “Highly positive correlations of muscle strength with lean mass in upper limbs
12 were found r-values 0.87-0.92 for boys and $r = 0.80-0.86$ for girls. High and moderate
13 positive correlations for lower limbs were also noted for upper limbs: $r = 0.74-0.86$ for
14 boys and $r = 0.67-0.82$ for girls.” (at 597). And, as observed by Zaras et al. (2020) in well
15 trained adult female weightlifters: “Very large to nearly perfect correlations were found
16 between snatch and clean and jerk for trunk lean body mass ($r = 0.959$ and 0.929) (at 1).”
17 The connection between lean body mass and muscle strength is quite clear, and the muscle
18 strength is very important to sports performance as stated by Comfort et al. (2023) in the
19 *National Strength and Conditioning Association Position Statement on Weightlifting for*
20 *Sports Performance*, “strength underpins performance in athletic tasks.” (at 1165)

21 36. In paragraph 19, Dr. Shumer cites the paper by Harper on Race Times for
22 Transgender Athletes as evidence that testosterone suppression and/or cross sex hormones
23 eliminates male advantages. Please see my report, paragraphs 155-159, for an explanation
24 of some of the numerous problems with the data from Harper. Also see my report
25 paragraphs 151-152 for analysis of the papers by Roberts et al. and Chicarelli et al.
26 regarding running times in transgender air force personnel, in which there is at least
27 objective evaluation of endurance performance in transwomen. Also see paragraph 169
28 for an explanation of the work by Alvares on VO₂max in transwomen.

1 37. In paragraphs 28-31, Dr. Shumer claims that the pre-pubertal male athletic
2 advantages are not due to “minipuberty”. At no point in my declaration are the male
3 athletic advantages differences ascribed to “minipuberty” (indeed, the term “minipuberty”
4 is not found within my expert report).

5 38. It is important to note that in their initial declarations, and in their rebuttal
6 statements, neither Dr. Stephanie Budge nor Dr. Daniel Shumer cited any peer reviewed
7 publications or presented any data demonstrating that the use of gonadotropin-releasing
8 hormone (GnRH) analogues (aka puberty blockers) prevent juvenile males from
9 developing male sex-based advantages in sports performance. Specifically, neither Dr.
10 Budge nor Dr. Shumer showed that the administration of puberty blockers causes males to
11 cease developing sex-based differences in lean body mass, body height, muscle strength,
12 muscle endurance, aerobic fitness, or any measure of sports-specific performance that gives
13 males large athletic advantages over comparably aged, gifted and trained females before
14 and after puberty.

15 39. In contrast, I presented considerable data and cited numerous peer reviewed
16 publications demonstrating that males have advantages in physical fitness and sports
17 performance before (see paragraphs 77-115, and the appendix, and the additional
18 information in my rebuttal to Dr. Shumer) and after puberty (see paragraphs 7-73). I also
19 cited and briefly summarized peer-reviewed publications demonstrating that administering
20 puberty blockers does not erase male sex-based advantages in lean body mass (see my
21 declaration paragraphs 117-121) and body height (see my declaration paragraphs 124 &
22 125).

23
24 I swear or affirm, under penalty of perjury, that the foregoing is true and correct.

25 Dated: June 29, 2023

/s/ Dr. Gregory A. Brown, Ph.D., FACSM

1 D. John Sauer, Mo. Bar No. 58721*
2 Justin D. Smith, Mo. Bar No. 63253*
3 James Otis Law Group, LLC
4 13321 North Outer Forty Road, Suite 300
5 St. Louis, Missouri 63017
6 Telephone: (314) 562-0031
7 John.Sauer@james-otis.com

8 *Attorneys for Intervenor-Defendants President Petersen and Speaker Toma*

9 **IN THE UNITED STATES DISTRICT COURT**
10 **FOR THE DISTRICT OF ARIZONA**
11 **TUCSON DIVISION**

12 Jane Doe, *et al.*,

13 Plaintiffs,

14 v.

15
16
17
18
19 Thomas C. Horne, in his official capacity
20 as State Superintendent of Public
21 Instruction, *et al.*,

22 Defendants.
23

Case No. 4:23-cv-00185-JGZ

**Rebuttal Declaration of James M.
Cantor, Ph.D., in Further Support of
Intervenor-Defendants' Opposition to
Plaintiffs' Motion for a Preliminary
Injunction**

24 I, James M. Cantor, declare as follows:

- 25 1. I have submitted an initial declaration to this Court dated May 18, 2023.
26 2. I now submit this expert rebuttal declaration based on my personal
27 knowledge, and it reflects my expert opinions.
28 3. In preparing this rebuttal declaration, I have reviewed the rebuttal

DEFENDANT'S
EXHIBIT

5

1 declarations filed by Plaintiffs, submitted by Dr. Shumer and Dr. Budge, and Plaintiffs’
 2 Reply Brief. I could not identify in Plaintiffs’ Reply any citation of my declaration or
 3 challenge to my claims.

4 **I. Response to Dr. Budge.**

5 4. Dr. Budge (Budge Rebuttal ¶6) claimed: “Dr. Cantor asserts that my claims
 6 about transgender youth and their medical care are based solely on my clinical experience.
 7 (Cantor Decl. ¶ 153.)” Her rebuttal claim is demonstrably untrue. As I made explicit in
 8 my declaration, I compared her claims with the research literature only to find she did not
 9 cite any of the research literature, so her claims did not come from it. (See Cantor Decl. ¶
 10 153)

11 5. Dr. Budge (Budge rebuttal ¶10) claimed: “Dr. Cantor further appears to
 12 suggest that excluding transgender girls from playing school sports is not psychologically
 13 damaging. (See Cantor Decl. ¶159.)” My actual words, however, indicate I said no such
 14 thing. Rather, I once again evaluated Dr. Budge’s claims using the standard criteria of the
 15 scientific method. Not only do I not suggest what is or is not damaging, but rather I am
 16 explicit in indicating that the lack of research does not allow anyone to know the answer
 17 to that inquiry. (See Cantor Decl. ¶ 159) In other words, when I clearly pointed out that it
 18 is not possible for Dr. Budge to know what she claims to know on this point, Dr. Budge
 19 interpreted this to mean that I somehow do know the opposite to be true, rather than my
 20 observation that no one can know.

21 6. Similarly, Dr. Budge (Budge rebuttal ¶9) inferred me to say the very opposite
 22 of what I did in my ¶ 158. She wrote: “Dr. Cantor appears to suggest that transgender girls
 23 can play on boys’ teams. (Cantor Decl. ¶ 158.)” My actual words, however, were
 24 unambiguous in referring to female-to-male students (whom Dr. Budge would call
 25 transgender boys) playing on male-designated teams: “The large majority of transgender
 26 adolescents are biologically female, and under SB-1165, continue to be permitted to
 27 participate on male designated teams, and these benefits remain available to them.”

28 7. Dr. Budge repeated her claims that the WPATH and Endocrine Society (E.S.)

standards are based on high quality reviews of the research. The only sources Dr. Budge cited in support of the claim of high quality were the E.S. and WPATH documents themselves. Dr. Budge did not, however, address the objective evidence: Table 1 of my declaration (Cantor decl pp. 31-32) lists the *entire set of every study* included in E.S. and WPATH reviews. The E.S. included exactly one study. WPATH included a total of four studies, *none* of which was a study of safety.

8. Dr. Budge's contention that "none of the international reports that Dr. Cantor cites is a clinical practice guideline" is frankly illogical: One does not write clinical guidelines for how to perform treatments *that are banned*. Indeed, Dr. Budge does not appreciate that this lack of clinical guidelines demonstrates that I am *correct* to refer to these policies as bans and that she is *incorrect* to represent the European policies as an attempt to increase the availability of medicalized transition. The purpose of the increase in clinics Dr. Budge mentioned, made very clear by NHS England, is to provide *mental health services* to these youth *instead of* easy access to medicalized transition. Dr. Budge did not refer to any of the other health care systems, referring only to English policy, calling it "an example." Dr. Budge did not contest, or even address, any of the statements made by any of the other national health care systems, all of which make incontrovertibly clear the intent of their policies were to ban medicalized transition.

9. Finally, the purpose of a systematic review is to provide the science, not decide the policy: Because my role as expert witness is to provide the science (and to let legislatures and courts decide policy), these reviews are exactly the appropriate documents for me to cite.

10. Dr. Budge's claims about "Dr. Cantor's view that..." (Budge rebuttal, header to her ¶18) and "Dr. Cantor appears to believe that banning..." (Budge rebuttal, ¶18), "Dr. Cantor appears to believe these minor patients..." (Budge rebuttal, ¶18) are objectionable for these same reasons. I never expressed these views, and they represent only more straw man arguments. Moreover, Dr. Budge's use of mental illness as a dismissive epithet reinforces the stigma of mental illness despite the goal of our shared profession to reduce

1 it.

2 11. In her original declaration, Dr. Budge (¶¶18–19) gave an invalid definition
3 of “sex,” incorrectly including “gender identity” as part of the definition of sex. To point
4 this out, my declaration provided the scientific definition of sex by including the verbatim
5 definitions from several professional associations in the U.S., all defining sex as I did, all
6 excluding gender identity from within the definition of sex, and all contradicting Dr.
7 Budge’s inclusion of gender identity within sex. In her rebuttal definition, Dr. Budge
8 claimed I used “outdated, inaccurate, and narrow definitions of sex” (Budge rebuttal ¶28)
9 and that the correct definitions were those provided by: “American Academy of Pediatrics,
10 2018; American Psychological Association, 2014; American Psychological Association,
11 2021; American Psychiatric Association, 2017; American Medical Association, 2018”
12 (Budge rebuttal ¶28).

13 12. The depth of Dr. Budge’s misapprehension is difficult to exaggerate: These
14 are largely the very documents I *did* cite, quoted verbatim, all showing they did *not* include
15 “gender identity.” For reference, the definition of sex in Dr. Budge’s original declaration
16 (¶19, italics added) was: “Every individual’s sex is comprised of many distinct biological
17 influences characteristics, including but not limited to chromosomal makeup hormones,
18 internal and external reproductive organs, secondary sex characteristics, *and gender*
19 *identity*.” My initial declaration (Cantor decl ¶¶106, 156) included these direct quotes:¹

20 a. Endocrine Society (Bhargava 2021 at 220.): “Sex is dichotomous,
21 with sex determination in the fertilized zygote stemming from unequal expression
22 of sex chromosomal genes.”

23 b. American Academy of Pediatrics (Rafferty 2018 at 2 Table 1.): “An
24 assignment that is made at birth, usually male or female, typically on the basis of
25 external genital anatomy but sometimes on the basis of internal gonads,
26 chromosomes, or hormone levels.”

27 c. American Psychological Association (APA Answers 2014): “Sex is

28 ¹ For reference, what Dr. Budge cited as AAP, 2018, I cited as Rafferty 2018.

1 assigned at birth, refers to one's biological status as either male or female, and is
2 associated primarily with physical attributes such as chromosomes, hormone
3 prevalence, and external and internal anatomy."

4 d. American Psychological Association (APA Resolution 2021 at 1):
5 "While gender refers to the trait characteristics and behaviors culturally associated
6 with one's sex assigned at birth, in some cases, gender may be distinct from the
7 physical markers of biological sex (e.g., genitals, chromosomes)."

8 e. American Psychiatric Association (Am. Psychiatric Ass'n Guide):
9 "Sex is often described as a biological construct defined on an anatomical,
10 hormonal, or genetic basis. In the U.S., individuals are assigned a sex at birth based
11 on external genitalia."

12 f. DSM-5-TR: In this chapter [on gender dysphoria], *sex* and *sexual*
13 refer to the biological indicators of male and female (understood in the context of
14 reproductive capacity), such as in sex chromosomes, gonads, sex hormones, and
15 nonambiguous internal and external genitalia. (American Psychiatric Association at
16 511, italics in original.)

17 13. None of the above definitions includes "gender identity," despite Dr.
18 Budge's now repeated insistence that they do. Dr. Budge included no evidence that any of
19 these documents contains what she claims they say.

20 14. Dr. Budge cited two documents not already quoted in mine, and they too fail
21 to support her definition of sex. The American Medical Association (2018) document was
22 a press release entitled "Affirming the Medical Spectrum of *Gender*" (italics added). The
23 quote Dr. Budge related was that "sex *and gender* are more complex," not that *sex* is more
24 complex. The press release indicated the need to understand "the scientific nature of a
25 person's sex" and refers to public policy implications following from legal definitions of
26 sex. The press release does not provide or cite any definition of sex of its own, as Dr.
27 Budge misrepresents.

28 15. Moreover, Dr. Budge's idiosyncratic terminology is not merely in error, and

1 not merely in conflict with her own cited sources, but it is also logically incoherent: If
 2 gender identity were an aspect included *within* the definition of sex, then it would not be
 3 possible for gender identity to be in conflict with sex, which is her definition of gender
 4 dysphoria. However much it might be Dr. Budge’s opinion that gender identity *should be*
 5 included, her own citations show it is not.

6 16. Where Dr. Budge claimed I said “sex can *only* be determined either by ‘visual
 7 inspection’ or ‘chromosome’” (Budge rebuttal declaration ¶28, italics added), I actually
 8 wrote “Sex is an objective feature: It can be ascertained regardless of any declaration by a
 9 person, *such as* by chromosomal analysis or visual inspection” (Cantor declaration ¶105,
 10 italics added). Whereas I am explicit in providing examples, Dr. Budge misrepresents it as
 11 an exhaustive list, ignoring entirely the key feature being their objectivity, not the examples
 12 of some objective features.

13 17. My initial declaration relayed the topics that WPATH and the Endocrine
 14 Society did and did not include in their research reviews, quoting directly those documents’
 15 reports of their own contents (Cantor decl ¶90 for Endocrine Society, ¶95 for WPATH).
 16 The Endocrine Society reviewed cross-sex hormones, but did not review puberty-blockers,
 17 and WPATH (published as Baker et al., 2021) did not review any safety studies, either of
 18 puberty-blockers or cross-sex hormones. WPATH’s review consisted of three studies
 19 (none of safety), and Endocrine Society’s, one. My declaration listed *every* study identified
 20 by WPATH and the Endocrine Society reviews (Cantor decl, Table 1, page 30): Neither
 21 Dr. Budge nor Dr. Shumer identified any study to be missing. Dr. Budge responded, not
 22 by citing any evidence for WPATH’s claims of safety or effectiveness, but by simply
 23 reiterating them (Budge rebuttal ¶36.) Dr. Budge claimed that these associations’
 24 guidelines were based on “comprehensive review of all existing data” (Budge rebuttal ¶
 25 36), and Dr. Shumer repeated that these associations’ guidelines are reliable “due to the
 26 rigorous nature of their review” (Shumer rebuttal ¶48), that they were “created based on
 27 rigorous reviews” (Shumer rebuttal ¶49), and “were developed through rigorous scientific
 28 processes” (Shumer rebuttal ¶49). Despite their liberal uses of these adjectives, neither Dr.

1 Budge nor Dr. Shumer presented any of the evidence that WPATH included to support
2 such claims.

3 18. Dr. Budge’s rebuttal (§37) and Dr. Shumer’s rebuttal (§50) erroneously
4 claimed it would not be ethical to conduct a Randomized Controlled Trial (RCT) to test
5 the effects of medicalized transition on minors. As already noted in my initial declaration
6 (section III.C.2.), when an experimental treatment cannot be compared with an *inactive*
7 treatment (a placebo), it can instead be compared with another *active* treatment (such as
8 psychotherapy), which is exactly what has been recommended by the England’s National
9 Health Service following its systematic review of effectiveness (Cantor decl §50). Neither
10 Drs. Budge nor Shumer contested or mentioned these facts.

11 19. Dr. Budge makes explicit that her belief that RCTs cannot be conducted is
12 because of the principle of “ equipoise,” which she defined as “true scientific uncertainty
13 about whether an intervention will help the individual.” (Budge rebuttal §37.) The idea is
14 that there must exist uncertainty in order to justify not providing the otherwise untested
15 treatment. Dr. Budge omitted an important part of the equipoise principle, however, failing
16 to spell out: uncertainty *among whom?* According to bioethicist Benjamin Freedman, the
17 originator of the concept of clinical equipoise, “The requirement is satisfied if there is
18 genuine uncertainty within *the expert medical community*—not necessarily on the part of
19 the individual investigator—about the preferred treatment.” (Freedman 1987².) That is, it
20 does not matter if Dr. Budge herself feels certain. The international expert medical
21 community is indeed highly *uncertain*, as already documented in my initial report (Section
22 II.F. ‘No Debate’) and to which neither Dr. Budge nor Dr. Shumer had a response. The
23 peer reviewed studies that have attempted to do so have been unable to demonstrate
24 differences in efficacy between medicalized and psychotherapeutic treatment of gender
25 dysphoria in minors. (See Section XIII. *Cohort Studies of Puberty Blockers and Cross-Sex*
26 *Hormones*.) That Dr. Budge and Dr. Shumer neglect to include disconfirming reports and

27 ² Freedman, B. (1987). Equipoise and the ethics of clinical research. *New England Journal*
28 *of Medicine*, 317, 141–145.

1 decline to acknowledge dissenting experts does not deny the widely documented existence
 2 of both. Indeed, the professional uncertainty is itself a topic covered by the international
 3 medical press, including the *British Medical Journal*'s recent article: *Gender Dysphoria in*
 4 *young people is rising and so is professional disagreement*. (Block 2023³.)

5 20. In her initial declaration, Dr. Budge asserted a long series of claims, none
 6 with any basis in science. She included no peer reviewed (or other) research at all to
 7 support any of these claims, as I indicated in my initial report. In her rebuttal declaration,
 8 Dr. Budge simply repeated her claim, "Studies have repeatedly documented that puberty
 9 blocking medication and hormone therapy are associated with mental health benefits for
 10 transgender people in both the short and long term, including a dramatically reduced rate
 11 of suicidality," this time including a footnote listing seven studies. (Budge rebuttal ¶38.)
 12 The Budge rebuttal included no summary, evaluation, or other indication of the contents
 13 of those studies.

14 21. For reference, the seven studies were: Tordoff 2022; Green 2021; Turban
 15 2020; Achille 2020; Kuper 2020; van der Miesen 2020; and Costa 2015. Of these seven,
 16 five are already detailed in my initial declaration, including the reasons they do not provide
 17 reliable evidence of benefit of medical transition (Cantor decl., section XIII.B. and
 18 XIII.C.), explicating either how they found medicalized transition *not* to be superior to
 19 psychotherapy (Costa 2015; Achille 2020; Tordoff 2022) or how they confounded medical
 20 transition with psychotherapy and so could not distinguish the source of any improvements
 21 (van der Miesen 2020; Kuper 2020). Dr. Budge did not contest any statement about these
 22 studies in my declaration, including no contest to my demonstration that these studies are
 23 scientifically unable to support the conclusions to which Dr. Budge puts them.

24 22. The remaining two of these seven studies are similarly unable to support Dr.
 25 Budge's conclusion. Both Green et al. (2021) and Turban et al. (2020) are surveys yielding
 26 correlational data. (Indeed, they are both from the same survey.) They do not demonstrate

27 ³ Block, J. (2023). Gender dysphoria in young people is rising—and so is professional
 28 disagreement. *British Medical Journal*, 380, p. 382.

1 that transition *caused* an improvement in mental health—They equally reflect that youth
2 with better mental health are the ones who receive permission to transition in the first place.
3 Dr. Budge did not contest my reminder that it is a scientific error to infer causation from
4 correlation. Yet she repeatedly engaged in that error.

5 23. Left off of Dr. Budge’s list of studies were the still other studies that
6 attempted to find improvement in transitioned groups but did not find such improvement.
7 (Cantor declaration, section XIII.A.) Dr. Budge’s rebuttal represents demonstrable cherry-
8 picking—citing only the studies that (seem to) suggest improvement while neglecting to
9 mention the others showing how unreliable the finding is. This is the very
10 misrepresentation of science that the systematic review process was designed to prevent.

11 24. Dr. Budge’s analysis of suicide and suicidality defy the scientific method.
12 As included in my initial declaration, “correlation does not imply causation.” Although
13 Dr. Budge did not contest that, her analysis of the suicide and suicidality data violated that
14 principle. Indeed, the Budge declaration makes explicit what her analysis was based on
15 (Budge rebuttal ¶45), and that basis is an explicit admission of doing exactly the opposite.
16 Dr. Budge’s exact words: “Suicide attempts and suicide are interrelated, a treatment that
17 reduces the former reduces the latter.” That mistaken belief is mistaken because it means
18 to infer causation from correlation.

19 **II. Response to Dr. Shumer.**

20 25. Dr. Shumer claimed my initial declaration “does not offer a single expert
21 opinion that directly relates to Arizona’s law” (Shumer rebuttal ¶42). Although the purpose
22 of my declaration is to assist the court rather than Dr. Shumer, the relevance of this material
23 is straight-forward: The perceived need to play on other-sex teams is to facilitate and
24 participate in the child’s social transition, which the evidence suggests greatly increases
25 the probability of medical transition. Thus, making an informed decision about beginning
26 that process requires the objective science for assessing claims about the outcomes of that
27 process.

28 26. Dr. Shumer’s initial declaration did not provide evidence-based opinions,

1 citing in total exactly one article, as I indicated in my initial declaration. In his rebuttal
 2 declaration, Dr. Shumer re-asserted his central claim, “Studies have repeatedly documented
 3 that pubertal suppression and hormone therapy are safe and effective treatments for
 4 transgender adolescents with gender dysphoria” (Shumer rebuttal ¶47), this time including
 5 a footnote listing eight studies. The list of eight is identical (and was even cited in the same
 6 order as) presented by Dr. Budge, with the exception of de Vries 2014 added to the end.
 7 That study was also discussed in my initial declaration, noting it was one of the studies to
 8 have confounded psychotherapy with medical transition.

9 27. Like Dr. Budge’s, Dr. Shumer’s rebuttal included no summary, evaluation,
 10 or other indication of evaluating the contents of those studies. Also like Dr. Budge, Dr.
 11 Shumer provided no indication that any of my descriptions of the contents of those studies
 12 was incorrect.

13 28. Dr. Shumer added that “These [eight] articles represent a small percentage
 14 of the full body of literature that was utilized to create evidence-based clinical practice
 15 guidelines for the treatment of gender dysphoria in children, adolescents, and adults”
 16 (Shumer rebuttal ¶47). First, Dr. Shumer’s addition of “*and adults*” hides from the reader
 17 just how little of the published evidence actually pertains to children and adolescents. The
 18 so-called “full body of literature” consists almost entirely of adults. Second, the set of
 19 studies included in my report is exhaustive—It includes all existing cohort studies reporting
 20 on the outcomes of these children. No study is left out, and Dr. Shumer cites no such study
 21 that I left out. The systematic reviews conducted by the multiple national public health
 22 care systems identified the same set of studies as appear in my report (with the exceptions
 23 of the most recent such studies, published afterwards). Dr. Shumer insinuates the existence
 24 of studies that do not, in fact, exist. Third, Dr. Shumer’s claim included that the “clinical
 25 practice guidelines” were based on “the full body of literature.” That claim also is untrue:
 26 Table 1 in my report (page 31) listed *every* study that was included in the systematic
 27 reviews from WPATH (e.g., Baker et al., 2021) and the Endocrine Society (Hembree et al.,
 28 2017). The Endocrine Society review included exactly *one* study about minors, and the

1 WPATH review included *zero* studies about safety. Studies cited outside the systematic
2 review process are subject to cherry-picking and other biases. There is no objective basis
3 justifying Dr. Shumer's claims.

4 29. My initial declaration presented the standard *Pyramid of Evidence* in science
5 (section III.A.) which illustrates the highest level of evidence being the systematic reviews
6 (section III.B.) and points out that randomized-controlled trials (RCTs) are the level
7 capable of distinguishing correlations (which provide only ambiguous conclusions) from
8 cause (section III.C.). In his rebuttal declaration, Dr. Shumer contradicts both himself and
9 the scientific method itself (Shumer rebuttal ¶50), writing: "Dr. Cantor also spends more
10 than 10 pages of his declaration discussing the 'Pyramid of Standards of Evidence' to
11 support his claim that the evidence supporting puberty suppression and hormone therapy
12 is not based on randomized controlled trials and is therefore not reliable. (Cantor Decl. ¶¶
13 38–66.) While I agree with Dr. Cantor that randomized control trials are an excellent study
14 design in many contexts, such trials are not ethically permissible for treatments that are
15 already known to provide a benefit to patients, which includes the use of GnRHa and
16 hormone therapy to treat gender dysphoria in adolescents."

17 30. First, Dr. Shumer's argument engages in a profound bait-and-switch: My
18 declaration (¶¶71–104) very clearly and consistently emphasized that it is the lack of
19 *systematic reviews* (the highest level) that are misleading many policies. As explicated in
20 my declaration, the national health care systems of several countries have conducted
21 systematic reviews (Cantor initial declaration, section VI.), but the U.S.-based clinical
22 groups have not (section, VII.).

23 31. Regarding the ethics of conducting RCTs, Dr. Shumer's bait-and-switch is
24 with *placebo-controlled* RCTs with *active-comparator* RCTs. As already noted (¶14
25 herein, regarding Dr. Budge's same error), an RCT with an active-comparator is the
26 medical standard and is recommended by England's National Health Service. Dr. Shumer
27 did not contest or mention these facts.

28 32. Third, Dr. Shumer is quite circular in his logic. He claims RCTs of transition

1 cannot be conducted because transition is “already known” to be beneficial; however, the
 2 only way to know what is beneficial is to conduct an RCT. As already noted in my initial
 3 report, it is not possible to infer causation from correlation (Cantor decl ¶56)—Dr. Shumer
 4 does not contest (or mention) this, but repeats exactly that error in claiming transition
 5 causes benefits on the basis of correlations only (i.e., evidence below the RCT-level on the
 6 *Pyramid of Evidence*). Rather than acknowledge current evidence to be limited and
 7 therefore able to justify only a cautious approach, Dr. Shumer is instead excusing the lack
 8 of strong evidence and advocating clinicians to proceed as if the evidence were definitive
 9 and unambiguous anyway.

10 33. Dr. Shumer did not contest, respond to, or mention my noting or the
 11 systematic reviews’ noting that, because none of the subset of studies suggesting possible
 12 improvement were RCTs, it cannot be known whether differences in mental health level
 13 were actually due to the psychotherapy the study participants were receiving at the same
 14 time, to placebo effects, to survivorship bias (regretful patients dropped out), selection bias
 15 (only healthier patients being permitted to transition in the first place), or other issues that
 16 require an RCT to be ruled out.

17 34. In his initial declaration, Dr. Shumer repeatedly claimed gender identity was
 18 *internal*—an *internal* sense (Shumer decl ¶23) and an *internal* phenomenon (Shumer decl
 19 ¶23). After the reminder in my declaration that science requires claims be objective,
 20 testable, and falsifiable (Cantor decl., IV.A.), Dr. Shumer’s rebuttal now claims instead
 21 that gender identity is “ascertained through *observation*” (Shumer rebuttal ¶59). Dr.
 22 Shumer is either attempting to have it both ways or does not know what these terms mean.
 23 If a feature is internal, it cannot be observed. If it can be observed, it is not internal.

24 35. There is little that can be said to Dr. Shumer’s profoundly anti-scientific
 25 belief that a concept being internal and ascertainable only by conversation “makes it not
 26 less valid or ‘scientific’” (Shumer rebuttal ¶59). In point of fact: Yes, it does. If being
 27 internal and ascertainable only by what someone says met the bar for “scientific,” then
 28 one’s telling of one’s internal religious experiences would represent scientific evidence for

1 the existence of divinity.

2 36. Dr. Shumer’s comments regarding attempts to “change” gender identity are
3 another bait-and-switch. The issue is not using psychotherapy to “change” gender identity,
4 but to help identity when and if a minor is mistaking their experiences to represent gender
5 dysphoria when it instead represents a different issue—one can be helped without the risks
6 posed by medicalized transition.

7 37. Dr. Shumer’s difficulties in accessing non-English research articles do not
8 change the international consensus of scientists. Moreover, Dr. Shumer is in error in
9 asserting that the Sweden research is unavailable in English: It is, in fact, published as a
10 peer-reviewed study in English (Ludvigsson, J. F., Adolfsson, J., Höistad, M., Rydelius,
11 P.-A., Kriström, B., & Landén, M. (2023). A systematic review of hormone treatment for
12 children with gender dysphoria and recommendations for research. *Acta Paediatrica*. doi:
13 10.1111/apa.16791).

14 38. Dr. Shumer has no basis for dismissing a claim I made because it was sourced
15 to a press release. My claim was about the status of the French public health care system,
16 and the press release was from that very ministry—it is a perfectly legitimate source for
17 that information. Dr. Shumer’s objection would be appropriate if it were the source for a
18 scientific fact, which, in my report, it was not.

19 39. In my initial report, I summarized the public health care systems of multiple
20 countries increasingly banning medicalized transition of minors, quoting their conclusions
21 directly. Dr. Shumer contested my use of the term “ban” as some of these policies permit
22 exceptions. Dr. Shumer’s analysis is in error: A “ban” in public health care is not the same
23 as a “ban” in the U.S. In a national, public health care system, it is possible to entertain
24 exceptions: All medical cases are known to the government, because all medical cases are
25 funded by the government. It is possible to know when only exceptional cases are being
26 treated as exceptions. The U.S. has no such centralized tracking system, supervision
27 system, or enforcement system. It is possible in these countries to detect when flexibility
28 is being abused, deeming nearly all transition-seekers to be “exceptions.” Because the U.S.

1 system has no such supervision or enforcement system, it includes the danger, arguably
2 even the history, of engaging in exactly such abuse.

3 40. Dr. Shumer is in error to cite Haldeman (2022) as evidence regarding
4 “conversion therapy” for gender identity. First, Haldeman (2022) is a collection of others’
5 writings, none of which is a peer-reviewed document. Second, Shumer engages in another
6 bait-and-switch, this time with sexual orientation with gender identity. The text below is
7 the overview published in Haldeman (2022). The overview includes both *sexual*
8 *orientation* change efforts (SOCE) and *gender identity* change efforts (GICE), and it
9 associates the ineffectiveness and only with SOCE, not GICE. Dr. Shumer, however,
10 applies it to GICE.

11
12
13 I swear or affirm, under penalty of perjury, that the foregoing is true and correct.

14 Dated: June 29, 2023

/s/ Dr. James M. Cantor, Ph.D.

1 D. John Sauer, Mo. Bar No. 58721*
2 Justin D. Smith, Mo. Bar No. 63253*
3 James Otis Law Group, LLC
4 13321 North Outer Forty Road, Suite 300
5 St. Louis, Missouri 63017
6 Telephone: (314) 562-0031
7 John.Sauer@james-otis.com

8 *Attorneys for Intervenor-Defendants President Petersen and Speaker Toma*

9 **IN THE UNITED STATES DISTRICT COURT**
10 **FOR THE DISTRICT OF ARIZONA**
11 **TUCSON DIVISION**

12 Jane Doe, *et al.*,

13 Plaintiffs,

14 v.

15
16
17
18
19 Thomas C. Horne, in his official capacity
20 as State Superintendent of Public
21 Instruction, *et al.*,

22 Defendants.
23

Case No. 4:23-cv-00185-JGZ

**Supplemental Declaration of Dr. Chad
Carlson, M.D., FACSM in Further
Support of Intervenor-Defendants'
Opposition to Plaintiffs' Motion for a
Preliminary Injunction**

24 I, Chad Carlson, declare as follows:

- 25 1. I have submitted an initial declaration to this Court dated May 18, 2023.
26 2. I now submit this supplemental declaration based on my personal knowledge,
27 and it reflects my expert opinions.
28 3. In preparing this supplemental declaration, I have reviewed the expert

DEFENDANT'S
EXHIBIT

6

1 rebuttal declarations filed by Plaintiffs, submitted by Dr. Shumer and Dr. Budge.

2 **I. There are pre-pubertal differences between boys and girls with respect to**
 3 **physical performance.**

4 4. Pre-pubertal boys and girls differ with respect to physical performance
 5 measures. This can in turn contribute to injury risk, as the modeling risk starts to rise when
 6 there are differences in strength, speed, and power between two groups. Thus, the
 7 performance side of this question is intricately linked to the safety side.

8 5. As has been pointed out, there are numerous cross-sectional studies from
 9 different countries looking at measured performance between the sexes in pre-adolescent
 10 children.¹ These studies generally show remarkable consistency, with boys outperforming
 11 girls in almost all age categories when looking at measures of speed, power and strength,
 12 and girls outperforming boys in flexibility assessments.

13 6. Plaintiffs' experts seek to dismiss these studies as not providing an accurate
 14 assessment of physiologic capacity, and instead reflecting social biases, because they are
 15 cross-sectional, meaning that the studies provide categorical comparison of aggregated
 16 data without attempting to assign causation to any measured differences, such as biology
 17 vs. sociological effect.

18 7. However, in his counter-argument, Dr. Shumer simply references a blog that
 19 in turn references another cross-sectional study to argue (in his words) that there is

21 ¹ See Catley M, et al. Normative Health-related Fitness Values for children: analysis of
 22 85347 test results on 9-17 year-old Australians since 1985. 47:98-108 (2013).
 23 DiMiguel-Etayo P., Physical fitness reference standards in European children: The
 24 IDEFICS Study. Int J Obes (Lond) 38(2):557-566 (2014).
 25 Kasovic M. et al., Secular trends in health-related physical fitness among 11-14 year old
 26 Croatian children and adolescents from 1999 to 2014. Scientific Rep 11:11039 (2021).
 27 Marta C. et al. Physical fitness differences between prepubescent boys and girls. J Strength
 28 Cond Res 26(7):1756-66 (2012).
 Tambalis K. et al., Physical fitness normative values for 6-18 year old Greek boys and
 girls. Eur J Sport Sci 16(6):736-46 (2016).
 Zhang F et al., Physical fitness reference standards for Chinese children and adolescents.
 Scientific Reports 11;4991 (2021).

1 supposedly “a well-established scientific consensus that, before puberty, there are no
2 significant differences in athletic performance between boys and girls.” The study linked
3 to in the blog is by McKay, in the Journal of Neurology.²

4 8. Although the McKay study does attempt to analyze data to discern the effect
5 of sex and body size, when you look at the actual study, as opposed to the blog summary,
6 Dr. Shumer’s claims cannot be justified.

7 9. First, the McKay study used a form of statistical analysis that is typically
8 used for comparison of continuous variables, not categorical variables such as sex, so from
9 the beginning, it is not well-suited to make inferences about differences in performance
10 between the sexes.³ That said, the McKay study looked at children aged 3-9 and found no
11 significant differences in strength measures between the sexes other than shoulder internal
12 rotation strength. (Notably, however, shoulder internal rotation strength is important for
13 throwing or striking a volleyball and probably a factor in early measured differences in
14 throwing capability between boys and girls (see below)).

15 10. From age 10, the McKay study aggregates data for analysis in a 10-19 year
16 old cohort, and there is a strong effect of sex on performance reported in that group. In the
17 10-19 year old cohort, boys were statistically stronger than girls in every measured strength
18 category. According to filed statements, plaintiffs went to at least Tanner II before they
19 started puberty suppression (approximately 11 years of age), so for purposes of this study,
20 they would have been aggregated into this group.

21 11. When Dr. Shumer says that “there was no statistical difference in the
22 capabilities of girls and boys until high-school age (commonly age 12),” what he appears
23 to be doing is repeating the summary conclusion noted on the referenced website, which in
24 turn appears to be using age 12 as a surrogate descriptor for the 10-19 year old group.
25 Unsurprisingly, aggregating ten and eleven year old boys with older male adolescents for

26
27 ² McKay MJ et al. Normative Reference Values for Strength and Flexibility of 1,000
Children and Adults. 88:36-43 (2017)

28 ³ Pearson Correlation - SPSS Tutorials - LibGuides at Kent State University

1 purposes of comparison to females does show a strong effect of sex on performance.

2 12. Thus, noting its previously referenced deficiencies regarding analysis of
3 performance differences between the sexes, the most one could really conclude from this
4 study is that there was no effect of sex on strength (other than shoulder internal rotation)
5 in the 3-9 year old cohort, which is not representative of the Plaintiffs (and which is not
6 what Dr. Shumer claimed anyway when he said “no significant differences”).

7 13. Returning to the previously referenced cross-sectional studies, the way that
8 these studies vary argues against Dr. Shumer’s interpretation. Measured differences exist
9 between the sexes when children are assessed in countries with varying social norms when
10 it comes to gender. Plaintiffs characterize the differences in these studies as “small,” but
11 in this context, “small” is a subjective term, and is used by Plaintiffs as framing language.

12 14. As an example, using Catley’s data on Australian children, when one looks
13 at performance at the 50th percentile for ten year old boys and girls, running times are
14 approximately 15% faster in boys in the 1.6 km run; boys complete double the number of
15 shuttle stages in a timed shuttle run, perform a standing long jump 7.5% farther, perform
16 30% more push-ups and demonstrate 10% greater handgrip strength as a group.⁴

17 15. Granted, these studies carry with them the limitations of cross-sectional
18 comparisons, but in terms of looking at measured trends across multiple studies, in diverse
19 regions with different social norms, there exists a consistency of outcome that is striking.

20 16. There is prospective data showing sex-based performance differences
21 between boys and girls. Dohrman looked at throwing and kicking ability of boys and girls
22 both before and after systematic skill coaching, and found that boys outperformed girls
23 both at baseline and after completion of training.⁵

24 17. Moreover, improvement was similar between the two groups. Differences
25 after training in throwing were nearly categorical.

26
27 ⁴ Catley, 2013.

28 ⁵ Dohrmann P. Throwing and kicking ability of 8 year old boys and girls. The Research
Quarterly 35(4):464-471 (1964).

1 18. One important thing to say about this study is that it occurred prior to Title
2 IX, when societal-based differences in sport-related skill sets between boys and girls would
3 be expected to be higher, and yet the rate of improvement between the two groups stayed
4 consistent. In other words, if girls were under-coached, providing them with equivalent
5 access to skill training should have resulted in greater improvement relative to boys; but
6 we don't see that.

7 19. Lombardo reviewed sex-based differences in throwing in children and found
8 similar consistent findings across all cultures. Boys throw farther in cultures where
9 overhead throwing sports do not predominate, and in cultures where overhead throwing is
10 common. These differences persist in cultures where boys and girls have similar
11 opportunities and would expect to gravitate toward one another in performance if social
12 bias were the primary determinant driving performance change. The following extended
13 quote from Lombardo et al. is instructive:

14
15 “At all ages, males throw faster, on average, than females. The sex
16 differences in throwing velocity emerge during the preschool years before
17 sex differences in sports involvement appear. By ages 4–7, 90% of boys
18 throw faster than the average of same-age girls (Thomas and French 1985)
19 and boys 6 years of age throw faster, on average, than do girls who are 9
20 years old (Rippe et al. 1990). Practice improves female throwing velocity
21 but, when compared to males with similar throwing experience, the sex
22 difference remains (Thomas and Marzke 1992; Petranek and Barton 2011).
23 By age 12, the fastest throwing girls are comparable to the slowest throwing
24 boys (Thomas and French 1985).

25
26 Throwing distance is positively and highly correlated with throwing speed
27 and is influenced by the point in the throw when the projectile is released
28 (Axe et al. 1996, 2014; Zhu et al. 2009). The magnitude of the sex differences

1 in the development of distance throwing exceeds that of any other motor skill
2 (Espenschade 1960) and manifests itself early; boys 6 years of age throw
3 farther than girls who are 9 years old (Rippe et al. 1990), and only the very
4 best girls throw as far as the least skilled boys at age 17 (Thomas and French
5 1985). Practice improves throwing distance for both sexes (Zhu et al. 2009),
6 but females do not “catch up” to males (Thomas et al. 1994). The inability of
7 girls to catch up to boys suggests that there may be a biological component
8 that limits the ability of females to improve their throwing with training
9 (Haubenstricker and Seefeldt 1986).

10
11 (Before adolescence) there is little sex difference in height, weight, and
12 muscle mass (National Center for Health Statistics 2017), so male superiority
13 in throwing cannot be attributed to sex differences in size or strength. Many
14 scholars of throwing attribute this developmental sex difference in throwing
15 to greater male participation in ball sports and social encouragement of
16 throwing skills by boys (Cratty 1979; Sakurai and Miyashita 1983; Bingham
17 and Souza 2009). However, Butterfield and Loovis (1993) showed that sex
18 differences in the throwing motion and performance remained after age and
19 Tee ball, softball, or baseball participation were statistically controlled
20 during analyses. These results are consistent with our hypothesis that the sex
21 difference in throwing may have a biological component....

22
23 Both male and female Australian Aborigines throw during hunting and
24 defense (Clarke 2003) leading Thomas et al. (2010) to predict that there
25 would be few sex differences in throwing in this population. However,
26 contrary to their prediction, boys still outperformed girls and the mean effect
27 sizes were large (Cohen 1992) but smaller than in same-aged populations in
28 the U.S...

1
2 Ehl et al. (2005) predicted that sex differences in throwing would be smaller
3 in Germany than in the U.S. where there are more sex differences in throwing
4 experience. They found sex differences in both populations and the relative
5 differences between the sexes were about the same. In terms of effect sizes,
6 and contrary to the authors' expectations, the differences between boys and
7 girls were larger in Germany ($d = 5.36$) than in the U.S. ($d = 1.82$). The
8 observation that German boys threw much better than girls in the U.S. even
9 though they do not throw much, have U.S.-like cultural support or
10 encouragement for throwing, and throw as well as U.S. boys (Ehl et al. 2005)
11 is consistent with our hypothesis and challenges the hypothesis that sex-
12 biased training significantly contributes to the sex differences in throwing.”⁶
13

14 20. All of this is noteworthy, because Dr. Shumer specifically states that
15 transgender girls who undergo early transition “do not have greater...throwing speed.”
16 Since throwing distance is dependent on both angle of release and initial throwing speed,
17 in light of the above data, including data from the very source cited by Dr. Shumer, his
18 contention cannot be true.

19 21. In addition to this, there is the question of physiologic ceiling. What is the
20 highest capability of boys and girls at different ages, in comparable competition? If there
21 is really very little difference between boys and girls physiologically, then prior to the
22 effects of testosterone occurring, or the pubertal growth spurt, records should be fairly
23 equally mixed between the sexes, at least in societies that promote more equal opportunities
24 for boys and girls to participate. But that is not what we see. USA Track and Field
25 maintains records by age category in boys' and girls' track. In nine representative events
26

27
28 ⁶ Lombardo MP et al. On the evolution of the sex differences in throwing: Throwing is a
male adaptation in humans. *The Quarterly Review of Biology* 93(2):91-119 (2018).

1 in the 8-and-under group,⁷ ten events in the 9-10 year old group,⁸ and fourteen events in
 2 the 11-12 year old group,⁹ boys' records exceed those of girls in 33/34 events. In the 11-
 3 12 year old long jump category, a girl holds the outright record, at 18' 1.5" compared to
 4 18' 1" for the boys. Even at a preadolescent age, the physiologic ceiling (represented by
 5 record performance) is different.

6 22. All this evidence suggests that there is, in fact, a biological effect of sex on
 7 sports performance before puberty, whether or not the cause of that effect has been fully
 8 elucidated yet. Available data makes a compelling case, and experience and common sense
 9 also testify to these differences.

10 11 **II. World Rugby and UK Sport.**

12 23. Plaintiffs' experts contend that World Rugby's current policy allows an
 13 exception for transgender women who transitioned "pre-puberty." But they misunderstand
 14 this point.

15 24. First, when you look at the referenced research that World Rugby used to
 16 come to their decision, it appears to entirely involve an older cohort. So their policy with
 17 regards to early transitioners is not grounded in scientific review of relevant data, at least
 18 based on what they have made available to the public.

19 25. Second, this exception is not an unqualified exception. The eligibility rules
 20 go on to stipulate that if an early transitioner wishes to participate in the female category,
 21 documentation of actual testosterone suppression must be provided. This is important,
 22 because many individuals on hormone suppression fail to suppress testosterone. It has
 23 been reported that this can occur in between 25% to 49% of treated subjects.¹⁰ If

24
25 ⁷ 100, 200, 400, 800, 1500, 4 x 100, 4 x 400, long jump, shot put.

26 ⁸ Same as above with addition of High Jump.

27 ⁹ Same as above with addition of 4 x 800, discus and mini-javelin.

28 ¹⁰ Heather AK. Transwomen elite athletes: Their extra percentage relative to female physiology. Int J Environ Res and Pub Health 19 (<https://doi.org/10.3390/ijerph19159103>) (2022).

1 testosterone drives major body changes in adolescence, it is certainly relevant whether
 2 medical treatments that are purporting to keep boys from entering male adolescence are
 3 actually working to achieve this goal in a given individual.

4 26. Third, national organizing bodies are already moving beyond these
 5 recommendations to further tighten controls. The Rugby Football Union in the UK now
 6 will not allow any exceptions for transgender females who did not start puberty suppression
 7 *before* age 11.¹¹ Thus, despite World Rugby’s recommendations about the preadolescent
 8 category, organizing unions under World Rugby are coming to their own conclusions.

9 27. Dr. Shumer criticizes inclusion of UK Sports’ published policy guideline on
 10 several points. First, he discounts this document because it “is not a scientific report.” This
 11 is interesting because he earlier tries to refute any idea of pre-adolescent performance
 12 advantage in boys by quoting a website. That said, even a cursory review of the
 13 background information that UK Sports reveals that they did, in fact, conduct an
 14 independent review of the available research and have made the references that they used
 15 to conduct this review publicly available.¹²

16 28. Unlike World Rugby, this review did consider some criteria involving
 17 younger subjects. To Dr. Shumer’s point, UK Sports goes on to state that “[c]urrent
 18 scientific evidence indicates that the difference between the strength, stamina, and
 19 physique between the sexes is *largely* due to the higher testosterone levels of males during
 20 their lifetime.” (Emphasis added.) This is not a controversial statement, and nobody is
 21 disputing it. But again, that does not mean that differences between the sexes before
 22 puberty don’t exist, or that they aren’t relevant to the conversation about public policy
 23 when it comes to sports.

24 **III. Dr. Shumer cannot support his assertion that puberty suppression and**
 25 **hormone replacement effectively eliminate the male advantage in sport.**

26 ¹¹ McLarnon M, et al. A Scoping Review of Transgender Policies in the 15 Most
 27 Commonly Played UK Professional Sports. Int J Environ Res Public Health. 2023 Feb
 28 17;20(4):3568.

¹² International Research Literature Review (equalityinsport.org).

29. As noted in my prior declaration, there is a dearth of studies directly addressing the question whether puberty suppression and hormone replacement effectively eliminate the male advantage in sports. Thus, we have to look at the evidence that is available and make inferences from that. However, it is important to point out that problems with lack of data cut against Plaintiffs here. Plaintiffs cannot point to any studies establishing that puberty suppression and hormone replacement eliminate the competitive advantage, as they must do to make their case. Policy makers who are trying to determine a way forward on issues pertaining to public safety, and in this case the integrity of sport, have to weigh what we do know. And, as stated above, considerations about pre-adolescent variation between the sexes do matter.

30. Dr. Shumer states that my declaration has “no relevance to transgender girls who transition early,” that “transgender girls who receive puberty suppressing medications at the onset of puberty... do not have any advantage over other girls with respect to size, weight, speed or strength,” and that I do “not cite to any evidence, nor does any exist, that such girls have an athletic advantage over other girls.” This is incorrect. In fact, I do cite studies that support this very conclusion.

31. One study I do cite to that is useful to consider in this regard is the Klaver study.¹³ This study looked at transgender girls placed on puberty suppression medication at various times, and tracked lean body mass (among other variables) prospectively. The study has been characterized as showing that biological males who transition end up with a biologically female body type. In fact, what this study does show is that males come into treatment with significantly higher lean body mass than their biological female peers, and that treatment in fact brings their lean body mass to an intermediate point that is still more than biological females are able to achieve. That contradicts the narrative constantly proffered by Dr. Shumer, that “transgender girls who receive puberty suppressing

¹³ Klaver M, de Mutsert R, Wiepjes CM, Twisk JWR, den Heijer M, Rotteveel J, Klink DT. Early Hormonal Treatment Affects Body Composition and Body Shape in Young Transgender Adolescents. J Sex Med. 2018 Feb;15(2):251-260. doi: 10.1016/j.jsxm.2017.12.009. PMID: 29425666.

1 medication at the onset of puberty do not differ from other girls with respect to the factors
2 that” I discuss in my declaration. At another point, Dr. Shumer notes that “a transgender
3 girl who received puberty suppressing medication followed by hormone replacement
4 therapy does not have the same physiology as a pre-pubertal, non-transgender boy.”
5 Perhaps, but such an individual is not a physiologic female either. In fact, when you look
6 at the data, the measured ranges for lean body mass in transgender girls fall above the mean
7 for cisgender women, indicating a categorical retained difference.

8 32. Klaver also tried to look at when individuals started therapy, and was not
9 able to show any differences in terms of anthropometric outcomes in individuals who
10 transitioned in early vs. later Tanner stages. That runs counter to what Dr. Shumer said in
11 his rebuttal declaration. It is also important to point out that these are not athletic subjects,
12 and that the effect of resistance training in an individual taking medication to suppress
13 puberty would be to mitigate the loss and maximize any retained difference between
14 transgender and cisgender girls.

15 **IV. Puberty suppression does not bring concussion risk to parity.**

16 33. Dr. Shumer appears to be missing the point on concussion risk. He states
17 that transgender girls will reduce their neck lean mass towards that of biological females,
18 bringing their risk more in-line with biological females. But the point is not that a
19 transgender girl might be increasing their injury risk by reducing lean body mass through
20 hormonal treatment. The point is that they are being placed into an environment where
21 they are now competing directly against biological girls who do carry higher susceptibility
22 for concussion injury. In other words, regardless of whether transgender girls become more
23 susceptible to concussions themselves, they are more likely to *cause* concussions than other
24 competitors when competing against biological girls.

25 34. Moreover, if the Klaver data is to be believed, transgender girls are
26 competing with a significant retained advantage in terms of lean body mass. Particularly
27 in a sport like volleyball, where, as we have noted above, boys have documented greater
28 arm strength and throwing speed, the ability to generate more ball force, in combination

1 with a biologically female population at higher susceptibility for head injury, places
 2 opposing female athletes at risk. Dr. Shumer states that “there is no evidence, and no
 3 medical reason to believe, that [the] participation [of transgender girls] on girls’ teams
 4 would pose any increased threat of...injuries to other girls.” But in fact, the evidence
 5 presents a compelling case that the opposite is true. It certainly is not misguided for policy
 6 makers to consider safety for all girls when deciding who should be allowed to compete
 7 together.

8 **V. Comments on two of the studies cited by Plaintiffs’ experts.**

9 A. *Harper, J. Race Times for Transgender Athletes. J Sporting Cultures and*
 10 *Identities 6(1) (2015).*

11 35. Were it not for the politically-charged subject matter, this study would never
 12 be published in most peer-reviewed journals. The whole point of the study is to compare
 13 race times of transgender female runners before and after transition, and assess the athlete’s
 14 placement relative to, first men, then women. Only eight runners were analyzed, and race
 15 times were self-reported and not uniformly verified. The author states in the methods
 16 section that “[f]or six of the eight runners, online checking made it possible to verify
 17 *approximately half of the submitted times*” (my emphasis). There was no assessment
 18 whether training across time remained consistent, or whether other factors than age may
 19 have mitigated performance.

20 36. Dr. Shumer’s closing sentence in paragraph 19 of his rebuttal, which deals
 21 with this study, has no bearing on the study itself.

22 B. *Tack LJW et al. Proandrogenic and Antiandrogenic progestins in*
 23 *transgender Youth: Differential Effects on Body Composition and Bone*
 24 *Metabolism. J Clin Endocrinol Metab 103(6):2147-2156.*

25 37. This study supports the conclusion that grip strength was retained, rather than
 26 Dr. Shumer’s characterization that subjects “did not experience any increase in grip
 27 strength.” This is relevant because ultimately the comparison is transwomen to cis-women,
 28 not cis-men. Dr. Shumer also notes that the study showed a significant decrease in lean

1 mass, but in reality, the results are mixed without any significant loss of lean muscle mass
2 in the legs. Again, the lesson here is that treatment does not turn a biological male into a
3 biological female. The natural inference is that there is retained difference.

4
5 I swear or affirm, under penalty of perjury, that the foregoing is true and correct.

6 Dated: June 29, 2023

/s/ Chad Carlson, M.D., FACSM

7 Chad Carlson, M.D., FACSM

8 Stadia Sports Medicine

9 West Des Moines, Iowa

10 Past President, AMSSM
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

Normative health-related fitness values for children: analysis of 85347 test results on 9–17-year-old Australians since 1985

Mark J Catley, Grant R Tomkinson

► The appendix to this paper is published online only. To view this file please visit the journal online (<http://bjsm.bmj.com/content/early/recent>)

Health and Use of Time (HUT) Group, Sansom Institute for Health Research, University of South Australia, Adelaide, Australia

Correspondence to

Dr Grant R Tomkinson, Health and Use of Time (HUT) Group, Sansom Institute for Health Research, University of South Australia, GPO Box 2471, Adelaide, SA 5001, Australia; grant.tomkinson@unisa.edu.au

Received 13 May 2011
Accepted 22 September 2011

ABSTRACT

Objectives To provide sex- and age-specific normative values for health-related fitness of 9–17-year-old Australians.

Methods A systematic literature search was undertaken to identify peer-reviewed studies reporting health-related fitness data on Australian children since 1985—the year of the last national fitness survey. Only data on reasonably representative samples of apparently healthy (free from known disease or injury) 9–17-year-old Australians, who were tested using field tests of health-related fitness, were included. Both raw and pseudo data (generated using Monte Carlo simulation) were combined with sex- and age-specific normative centile values generated using the Lambda Mu and Sigma (LMS) method. Sex- and age-related differences were expressed as standardised effect sizes.

Results Normative values were displayed as tabulated percentiles and as smoothed centile curves for nine health-related fitness tests based on a dataset comprising 85347 test performances. Boys typically scored higher than girls on cardiovascular endurance, muscular strength, muscular endurance, speed and power tests, but lower on the flexibility test. The magnitude of the age-related changes was generally larger for boys than for girls, especially during the teenage years.

Conclusion This study provides the most up-to-date sex- and age-specific normative centile values for the health-related fitness of Australian children that can be used as benchmark values for health and fitness screening and surveillance systems.

BACKGROUND

Physical fitness is considered to be an important marker of current and future health in children and adults.¹ In children, cardiovascular fitness is a weak-to-strong predictor of total and abdominal adiposity, cardiovascular disease risk factors, cancer and mental health.^{1–2} Certain muscular fitness components (eg, strength and endurance) are moderate predictors of cardiovascular disease risk factors, skeletal health and mental health.¹ Meaningful relationships have also been reported between running speed (another muscular fitness component) and skeletal health.³ In adults, cardiovascular fitness is a strong and independent predictor of all-cause mortality and cardiovascular disease mortality and morbidity,⁴ stroke,⁵ cancer, mental health,⁶ health-related quality of life⁷ and many other cardiometabolic risk factors and comorbidities.^{8–9} Moreover, physical fitness

tracks moderately well from childhood through to adulthood.^{10–13} This evidence highlights the need to include health-related fitness testing (ie, the testing of fitness components such as cardiovascular and muscular fitness that have the strongest links with health outcomes) as part of existing health and fitness screening and surveillance systems.

Although the most valid assessments of fitness require sophisticated laboratory equipment and a high level of tester expertise, they unfortunately are not suitable for mass testing. On the other hand, properly conducted field tests offer simple, feasible, and practical alternatives, which typically demonstrate good reliability and validity.^{2–14–17} In Australia, unlike in Europe and North America where standardised test batteries such as the Eurofit¹⁸ or FITNESSGRAM¹⁹ are widely administered, a number of different field-based fitness tests and testing protocols have been used over time. For example, the most popular field test of cardiovascular fitness in Australia in the 1960s and 1970s was the 549-m (600 yd) run; in the 1980s and 1990s, it was the 1600-m run; and over the past decade or so, it has been the 20-m shuttle run.²⁰ Many physical educators and sports coaches in Australia continue to administer tests that are no longer in favour, largely because normative data (which are now several decades old) are available. This makes it difficult to assess the current status of health-related fitness in Australian children. Further compounding the problem is that the last national fitness survey of Australian children was conducted in 1985,²¹ and with convincing evidence of recent temporal changes in several components of fitness,^{22–24} the usefulness of such data seems to be limited.

Because there has never been a follow-up to the 1985 national survey, this study aimed to locate large and reasonably representative datasets of Australian children to generate normative centile values for health-related fitness. This study also aimed to quantify sex- and age-related differences in health-related fitness. These normative data will facilitate the identification of children with (a) low fitness in order to set appropriate goals and to promote positive health behaviours, and (b) specific fitness characteristics that may be considered important for sporting success.

METHODS

Data sources

A systematic review of the peer-reviewed scientific literature was undertaken to locate studies



Original Article

Table 1 Summary of the included studies that have been used to assess the health-related fitness of 9–17-year-old Australians since 1985

Study	Year	Age (years)	N	Raw data	Sampling method	Sample base	Protocol	Tests reported in included studies								
								Push-ups	Sit-ups	Standing broad jump	Basketball throw	50 m sprint	Sit-and-reach	Hand-grip	1.6 km run	20 m shuttle run
ACHPER ⁵⁰	1994	9–18	39–104	yes	School-based; stratified, proportional	State (VIC)	ACHPER ⁵⁰	*	*	*	*	*	*	*	*	*
Barnett <i>et al</i> ⁵¹	2007	15–17	21–69	no	School-based; stratified, random	State (NSW)	ACHPER ⁵⁰									*
Birchall ⁵²	1990	5–12	6–184	yes	School-based; convenience	State (VIC)	Pyke ²¹	*	*	*		*	*		*	
Booth <i>et al</i> ⁵³	1997	9, 11, 13, 15	399–634	no	School-based; stratified, proportional	State (NSW)	ACHPER ⁵⁰		*		*					*
Booth <i>et al</i> ⁵⁴	2004	9–15	357–466	no	School-based; stratified, proportional	State (NSW)	ACHPER ⁵⁰									*
Burke <i>et al</i> ⁵⁵	2004	10–13	38–117	yes	School-based; stratified, proportional	Capital city (WA)	ACHPER ⁵⁰									*
Cooley and McNaughton ⁵⁵	1998	11–16	339–636	no	School-based; stratified, proportional	State (TAS)	ACHPER ⁵⁰									*
Dollman <i>et al</i> ⁵⁷	1997	10–12	118–450	yes	School-based; stratified, proportional	State (SA)	Pyke ²¹			*		*			*	
Dollman pers. comm.	2002	11–12	19–154	yes	School-based; stratified, random	State (SA)	Pyke ²¹						*	*		
Dollman, pers. comm.	2002	8–12	8–389	yes	School-based; stratified, proportional	State (SA)	ACHPER ⁵⁰ Pyke ²¹			*						*
Hands ⁵⁸	2000	6–12	14–37	yes	School-based; stratified, random	Capital city (WA)	ACHPER ⁵⁰ Pyke ²¹	*	*	*	*	*	*	*	*	*
McIntyre, pers. comm.	2009	10–11	23–44	yes	School-based; stratified, random	Capital city (WA)	ACHPER ⁵⁰									*
McNaughton <i>et al</i> ⁵⁹	1995	7–10	30–83	no	School-based; stratified, random	State (TAS)	Pyke ²¹					*			*	
Pyke ²¹	1985	7–15	405–497	yes	School-based; stratified, proportional	National	Pyke ²¹	*	*	*		*	*	*	*	*
Vandongen <i>et al</i> ⁶⁰	1990	11	485–486	no	School-based; stratified, random	Capital city (WA)	ACHPER ⁵⁰								*	*

identifies test data that are available.

ACHPER, Australian Council for Health, Physical Education and Recreation; year, year of testing; n, sample size range per sex by age by test group; VIC, TAS, SA, WA, NSW

reporting descriptive summary data on Australian children tested for health-related fitness using field tests. Candidate studies were searched for in November 2009 using a computer search of online bibliographic databases (Ausport, CINAHL, Medline, PubMed, Scopus and SPORTDiscus). The search string used for the computer search was: (((((((((((fitness) OR aerobic) OR anaerobic) OR cardio*) OR endurance) OR agility) OR flexibility) OR speed) OR power) OR strength) OR sprint*) OR jump*) OR push-up*) OR sit-up*) OR grip strength) OR sit and reach) AND (((((((child*) OR paediatric*) OR adolesc*) OR boy*) OR girl*) OR youth*) OR teen*) AND (Australia*). All titles and abstracts (when available) were assessed to identify eligible articles, with full-text articles retrieved if there was doubt in an article's eligibility. A number of Australian researchers were contacted through email

to ask whether they knew of any appropriate studies or unpublished datasets.

Inclusion/exclusion criteria

Studies were included if they explicitly reported descriptive health-related fitness test data for apparently healthy (free from known disease or injury) 9–17-year-old Australians who were tested from 1985 onwards and if they reported data at the sex by age by test level, on children directly measured using field-based fitness tests for which explicit testing protocols were available. Studies were excluded if they reported descriptive data that were published in another identified study. The reference lists of all included studies were examined and cross-referenced to identify additional studies. Attempts were made to contact the corresponding

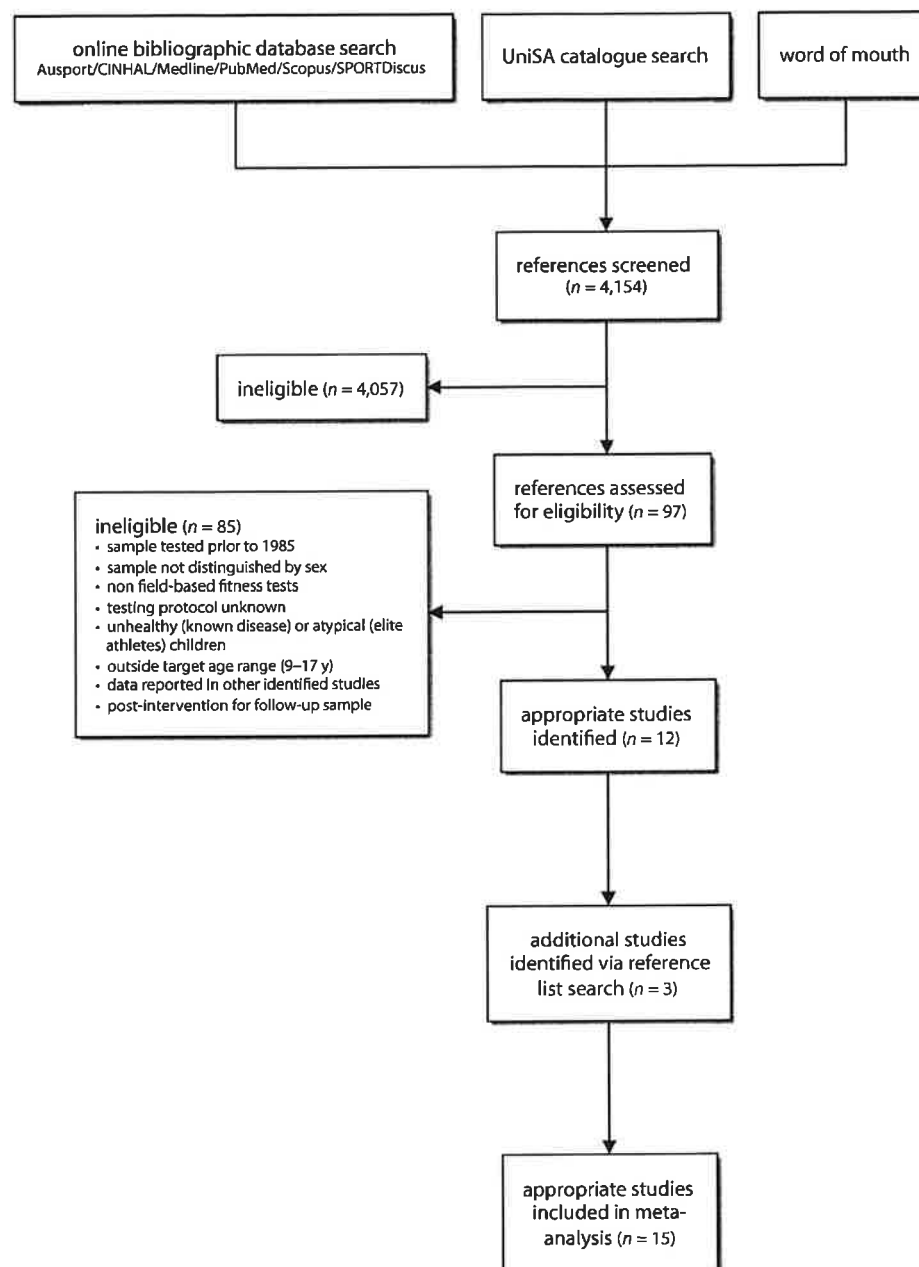


Figure 1 Flowchart outlining the identification of the included studies.

author of each study to request raw data and/or to clarify study details.

Initial data analysis

The following descriptive data were extracted from each included study: sex, age, year of testing, sample size, mean, SD, fitness test type and test protocol. Only data for commonly used fitness tests that were collected using protocols that were originally described in national or state-based fitness surveys of Australian children were retained for further analysis. Tests were considered to be 'common' if they were used to measure fitness in children across a broad range of ages and in at least two separate studies. Data for each fitness test were expressed in a common metric, and protocol differences were corrected where possible (eg, 20 m shuttle

run data were expressed as the number of completed stages using the correction procedures described by Tomkinson *et al*).²⁵ However, if protocol correction was not possible, then only fitness data collected using the most common test protocol were retained. All available raw data were checked for anomalies by running range checks with data ± 3 SDs away from the respective study by sex by age by test mean excluded. Age was expressed in whole years as the age at last birthday.

Statistical analysis

Sex- and age-specific normative centile values were calculated using a dataset comprising raw data and pseudo data that were generated using the method described by Tomkinson.²⁴ Normative centile values were generated using

Original Article

Table 2 1.6 km run (s) centile values and LMS summary statistics by sex and age in 9- to 17-year-old Australians

Age (year)	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅	L	M	S
Boys														
9	750	684	618	578	547	522	499	476	452	423	401	-1.042	521.963	0.183
10	732	666	602	564	535	511	489	469	447	420	400	-1.284	511.053	0.175
11	710	646	585	549	523	500	480	461	441	416	397	-1.466	500.394	0.166
12	682	621	563	530	505	485	467	449	430	408	392	-1.721	484.819	0.157
13	643	587	535	505	483	465	448	432	415	395	380	-1.895	464.529	0.148
14	605	556	509	482	462	446	431	416	401	382	369	-1.987	445.569	0.140
15	575	531	490	465	447	432	418	404	390	373	360	-1.979	431.504	0.133
16	552	514	477	454	437	423	410	397	383	366	354	-1.865	422.693	0.128
17	534	500	467	446	430	417	404	392	379	362	350	-1.707	416.545	0.123
Girls														
9	829	769	706	666	635	609	584	559	533	499	475	-0.779	608.674	0.167
10	820	759	697	657	626	600	576	552	526	494	470	-0.878	600.149	0.166
11	801	741	680	641	611	586	562	539	514	483	460	-0.929	585.820	0.165
12	784	726	666	629	600	575	552	529	505	474	452	-0.921	574.682	0.164
13	771	716	658	621	593	569	546	524	500	469	447	-0.852	568.706	0.163
14	763	711	655	620	592	567	545	523	498	468	445	-0.737	567.486	0.162
15	760	710	656	621	594	570	547	525	500	469	446	-0.591	569.809	0.161
16	757	710	658	624	597	573	550	527	502	471	446	-0.428	572.723	0.160
17	753	708	658	625	598	575	552	529	504	471	446	-0.263	574.536	0.159

Note, percentile data were calculated from 11 423 1.6 km run performances collected between 1985 and 1997.
L, skew; M, median; P, percentile; S, coefficient of variation.

Table 3 20 m shuttle run (completed stages) centile values and LMS summary statistics by sex and age in 9- to 17-year-old Australians

Age (year)	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅	L	M	S
Boys														
9	1	1	2	2	2	3	3	3	4	5	6	0.213	2.573	0.568
10	1	2	2	3	3	4	4	5	5	7	8	0.373	3.537	0.543
11	1	2	3	3	4	4	5	5	6	7	8	0.520	4.131	0.517
12	1	2	3	3	4	4	5	6	6	8	9	0.643	4.460	0.486
13	2	2	3	4	4	5	5	6	7	8	9	0.744	4.888	0.453
14	2	3	4	4	5	6	6	7	8	9	10	0.835	5.664	0.418
15	3	3	4	5	6	7	7	8	9	10	11	0.926	6.527	0.380
16	3	4	5	6	7	7	8	8	9	10	11	1.031	7.159	0.343
17	4	5	6	6	7	8	8	9	10	11	11	1.143	7.690	0.306
Girls														
9	1	1	1	1	2	2	2	3	4	5	5	-0.065	1.842	0.535
10	1	1	2	2	2	2	3	3	4	5	6	0.086	2.468	0.557
11	1	1	2	2	2	3	3	4	4	6	7	0.220	2.844	0.573
12	1	1	2	2	3	3	3	4	5	6	7	0.324	3.016	0.577
13	1	1	2	2	3	3	4	4	5	6	7	0.400	3.138	0.569
14	1	1	2	2	3	3	4	4	5	6	7	0.457	3.225	0.554
15	1	1	2	3	3	3	4	4	5	6	7	0.505	3.412	0.536
16	1	2	2	3	3	4	4	5	5	6	7	0.554	3.672	0.518
17	1	2	2	3	4	4	5	5	6	7	8	0.603	4.032	0.499

Percentile data were calculated from 18 075 20 m shuttle run performances collected between 1990 and 2009.

The 20 m shuttle run can be scored in different metrics other than as the number of completed stages, such as the number of completed laps, the speed at the last completed stage and as mass-specific peak oxygen uptake estimates (see Tomkinson *et al*²⁵ for details on how to correct 20 m shuttle run performances to different metrics).
L, skew; M, median; P, percentile; S, coefficient of variation.

LMSChartmaker Light (v2.3, The Institute of Child Health, London) software, which analyses data using the LMS method.²⁶ The LMS method fits smooth centile curves to reference data by summarising the changing distribution of three sex- and age-specific curves representing the skewness (L: expressed as a Box-Cox power), the median (M) and the coefficient of variation (S). Using penalised likelihood, the curves can be fitted as cubic splines using non-linear regression, and

the extent of smoothing required can be expressed in terms of smoothing parameters or equivalent degrees of freedom.²⁷

For each fitness test, differences in means between: (a) age-matched Australian boys and girls (eg, 10-year-old boys vs 10-year-old girls); (b) sex-matched Australian children of different ages (eg, 10-year-old boys vs 11-year-old boys); and (c) sex- and age-matched Australian and international children^{18 28–30} were expressed as standardised effect sizes.³¹ Positive effect

Table 4 50 m sprint (s) centile values and LMS summary statistics by sex and age in 9- to 15-year-old Australians

Age (year)	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅	L	M	S
Boys														
9	10.6	10.2	9.8	9.5	9.3	9.1	9.0	8.8	8.6	8.3	8.1	-1.837	9.136	0.078
10	10.5	10.1	9.7	9.4	9.2	9.0	8.8	8.7	8.5	8.2	8.0	-2.185	9.009	0.080
11	10.4	10.0	9.6	9.3	9.1	8.9	8.7	8.5	8.3	8.1	7.9	-2.405	8.877	0.081
12	10.2	9.8	9.3	9.1	8.9	8.7	8.5	8.3	8.1	7.9	7.7	-2.446	8.673	0.081
13	9.8	9.4	9.0	8.8	8.6	8.4	8.2	8.1	7.9	7.7	7.5	-2.489	8.377	0.079
14	9.4	9.0	8.7	8.4	8.2	8.1	7.9	7.8	7.6	7.4	7.2	-2.701	8.063	0.076
15	9.0	8.6	8.3	8.1	7.9	7.7	7.6	7.5	7.3	7.1	7.0	-3.021	7.738	0.073
Girls														
9	11.7	11.3	10.8	10.5	10.3	10.0	9.8	9.6	9.3	9.0	8.8	-0.981	10.033	0.088
10	11.1	10.7	10.3	10.0	9.8	9.5	9.3	9.1	8.9	8.6	8.4	-1.453	9.542	0.084
11	10.7	10.3	9.9	9.6	9.4	9.2	9.0	8.8	8.6	8.3	8.1	-1.803	9.161	0.082
12	10.4	10.0	9.6	9.3	9.1	8.9	8.7	8.6	8.4	8.1	7.9	-1.977	8.919	0.080
13	10.2	9.8	9.4	9.2	9.0	8.8	8.6	8.4	8.3	8.0	7.8	-1.991	8.787	0.078
14	10.0	9.7	9.3	9.1	8.9	8.7	8.5	8.4	8.2	7.9	7.8	-1.884	8.686	0.076
15	9.9	9.6	9.2	9.0	8.8	8.6	8.5	8.3	8.1	7.9	7.7	-1.724	8.638	0.075

Note, percentile data were calculated from 10 104 50 m sprint performances collected between 1985 and 1999. L, skew; M, median; P, percentile; S, coefficient of variation.

Table 5 Basketball throw (m) centile values and LMS summary statistics by sex and age in 9- to 17-year-old Australians

Age (year)	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅	L	M	S
Boys														
9	2.3	2.5	2.7	2.9	3.1	3.3	3.4	3.6	3.8	4.1	4.4	0.623	3.260	0.198
10	2.5	2.8	3.0	3.3	3.4	3.6	3.8	4.0	4.2	4.5	4.8	0.675	3.608	0.192
11	2.8	3.1	3.4	3.6	3.8	4.0	4.2	4.4	4.7	5.0	5.3	0.733	4.026	0.188
12	3.1	3.4	3.8	4.0	4.3	4.5	4.7	4.9	5.2	5.6	5.9	0.792	4.471	0.188
13	3.5	3.8	4.2	4.5	4.8	5.0	5.3	5.5	5.8	6.2	6.6	0.843	5.012	0.187
14	3.9	4.2	4.7	5.0	5.3	5.5	5.8	6.1	6.4	6.9	7.2	0.898	5.522	0.186
15	4.2	4.6	5.0	5.4	5.7	6.0	6.3	6.6	6.9	7.4	7.8	0.943	5.975	0.185
16	4.4	4.8	5.3	5.6	6.0	6.3	6.5	6.9	7.2	7.7	8.2	0.964	6.254	0.185
17	4.5	4.9	5.5	5.8	6.2	6.5	6.8	7.1	7.5	8.0	8.5	0.966	6.467	0.187
Girls														
9	2.1	2.3	2.5	2.7	2.9	3.0	3.2	3.3	3.5	3.7	3.9	1.116	3.015	0.182
10	2.3	2.6	2.8	3.0	3.2	3.3	3.5	3.7	3.8	4.1	4.3	1.024	3.336	0.181
11	2.6	2.8	3.1	3.3	3.5	3.6	3.8	4.0	4.2	4.5	4.7	0.942	3.646	0.180
12	2.8	3.1	3.4	3.6	3.8	4.0	4.2	4.3	4.6	4.9	5.2	0.873	3.970	0.180
13	3.0	3.3	3.6	3.9	4.1	4.3	4.5	4.7	4.9	5.3	5.6	0.816	4.265	0.179
14	3.2	3.4	3.8	4.0	4.2	4.4	4.6	4.8	5.1	5.4	5.7	0.739	4.410	0.175
15	3.3	3.6	3.9	4.1	4.3	4.5	4.7	4.9	5.1	5.5	5.8	0.606	4.486	0.169
16	3.4	3.7	4.0	4.2	4.4	4.6	4.7	5.0	5.2	5.6	5.9	0.394	4.557	0.162
17	3.6	3.8	4.1	4.3	4.5	4.6	4.8	5.0	5.3	5.6	5.9	0.140	4.634	0.154

Note, percentile data were calculated from 5,541 basketball throw performances collected between 1994 and 1999; L, skew; M, median; P, percentile; S, coefficient of variation.

sizes indicated that mean fitness test performances for boys (age-matched analysis), older children (sex-matched analysis) or Australian children (sex- and age-matched analysis) were higher than those for girls, younger children or international children, respectively. Effect sizes of 0.2, 0.5 and 0.8 were used as thresholds for small, moderate and large.³¹

RESULTS

Table 1 summarises the 15 included studies. Of these, 12 were identified through bibliographic database searching and word of mouth, and three were identified through reference list searching. Corresponding authors of all the studies were

contacted through email to clarify study details and/or to request raw data. All authors satisfactorily clarified study details, and seven of them supplied raw data (figure 1).

The final dataset comprised 85347 individual test results and 142 sex by age by test groups with a median sample size of 537 (range: 54–2612). Data were available for six fitness components and nine fitness tests: cardiovascular endurance (20 m shuttle run, 1.6 km run), muscular strength (hand-grip), muscular endurance (push-ups and sit-ups), muscular power (standing broad jump and basketball throw), muscular speed (50 m sprint) and flexibility (sit-and-reach). Raw data were available for 74% of all data points.

Original Article

Table 6 Standing broad jump (cm) centile values and LMS summary statistics by sex and age in 9- to 15-year-old Australians

Age (year)	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅	L	M	S
Boys														
9	105	113	121	127	133	138	142	147	153	161	168	1.244	137.506	0.138
10	109	117	126	133	138	143	148	154	160	168	174	1.490	143.430	0.138
11	112	121	131	138	144	149	154	160	166	174	181	1.654	149.322	0.138
12	117	126	137	144	150	156	161	167	173	182	189	1.704	155.838	0.137
13	126	136	147	154	161	166	172	178	185	194	201	1.629	166.340	0.135
14	137	146	157	165	172	178	184	190	197	206	214	1.526	177.688	0.131
15	148	157	169	177	183	189	196	202	209	219	228	1.446	189.485	0.127
Girls														
9	95	102	110	116	122	126	131	136	142	150	157	1.098	126.379	0.149
10	100	108	117	123	128	133	138	143	149	158	165	1.152	133.177	0.147
11	106	114	123	129	135	140	145	151	157	166	173	1.197	140.142	0.145
12	111	118	128	135	140	145	151	156	163	171	179	1.211	145.432	0.142
13	115	123	132	139	145	150	155	161	167	176	183	1.183	150.080	0.138
14	119	127	136	143	148	154	159	164	171	180	187	1.158	153.551	0.134
15	122	129	139	145	151	156	161	166	173	181	188	1.148	155.661	0.130

Percentile data were calculated from 11 194 standing broad jump performances collected between 1985 and 2002.
L, skew; M, median; P, percentile; S, coefficient of variation.

Table 7 Push-ups (no. in 30 s) centile values and LMS summary statistics by sex and age in 9- to 15-year-old Australians

Age (year)	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅	L	M	S
Boys														
9	4	6	8	9	11	12	14	15	17	20	22	0.846	12.310	0.452
10	4	6	8	10	11	13	14	16	18	21	23	0.894	12.943	0.447
11	4	6	8	10	12	13	14	16	18	20	22	0.940	12.942	0.438
12	4	6	9	10	12	13	15	16	18	20	22	0.980	13.200	0.422
13	5	7	9	11	13	14	16	17	19	22	24	1.020	14.255	0.399
14	6	8	11	13	14	16	17	19	21	23	25	1.070	15.954	0.370
15	7	10	13	15	16	18	19	21	23	25	27	1.126	17.697	0.337
Girls														
9	2	3	5	7	8	9	10	12	13	16	18	0.719	8.989	0.550
10	2	3	5	6	7	9	10	11	13	16	18	0.652	8.655	0.583
11	2	3	4	6	7	8	9	11	13	16	18	0.584	8.142	0.624
12	1	2	4	5	6	7	9	10	12	15	18	0.518	7.395	0.672
13	1	2	3	4	6	7	8	10	12	15	18	0.453	6.792	0.720
14	1	2	3	4	5	6	8	9	11	15	18	0.390	6.384	0.765
15	1	2	3	4	5	6	7	9	11	14	18	0.329	5.818	0.812

Percentile data were calculated from 7,342 push-up test performances collected between 1985 and 1991.
L, skew; M, median; P, percentile; S, coefficient of variation.

Normative fitness data for 9–17-year-old Australians are presented as tabulated percentiles from 5 to 95 (P₅, P₁₀, P₂₀, P₃₀, P₄₀, P₅₀, P₆₀, P₇₀, P₈₀, P₉₀, P₉₅) in tables 2–10. The sex- and age-specific LMS values for all fitness tests are also shown. The LMS values depict the nature of the age-related distributions for boys and girls and can be used to calculate z-scores and hence percentile values by looking up a z-table, using the following formula:

$$z = \frac{\left(\frac{x}{M}\right)^L - 1}{L \times S}$$

where z is z score, x is performance, L is sex- and age-specific L-value, M is the sex- and age-specific M-value and S is the sex- and age-specific S-value.

Figures 2 and 3 show the smoothed centile curves (P₁₀, P₅₀, P₉₀).

Figure 4 shows the sex-related differences in mean fitness. Boys consistently scored higher than girls on health-related fitness tests, except on the sit-and-reach test, with the magnitude of the differences typically increasing with age and often accelerating from about 12 years of age. Overall, the magnitude of differences between boys and girls was large for the 1.6 km run, 20 m shuttle run, basketball throw and push-ups; moderate for the 50-m sprint, standing broad jump and sit-and-reach; and small for sit-ups and hand-grip strength. Figure 5 shows the age-related changes in mean fitness. The age-related changes were typically larger for boys than for girls, especially during the teenage years, and for muscular fitness tests than for cardiovascular fitness tests. Fitness also tended to peak from about the age of 15 years. Figure 6 shows that the differences in health-related fitness between Australian and international children

Table 8 Sit-ups (no. in 180 s) centile values and LMS summary statistics by sex and age in 9- to 17-year-old Australians

Age (year)	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅	L	M	S
Boys														
9	3	5	8	11	14	17	21	25	30	40	48	0.321	17.046	0.755
10	5	8	13	17	20	24	29	34	40	50	59	0.466	24.459	0.669
11	6	10	16	21	25	29	34	39	45	55	60	0.629	29.422	0.594
12	8	14	21	26	31	36	40	45	51	60	60	0.841	35.561	0.514
13	10	17	25	31	36	40	45	50	55	60	60	1.056	40.288	0.443
14	12	20	29	34	39	43	48	52	57	60	60	1.232	43.454	0.389
15	14	22	31	36	41	45	49	53	58	60	60	1.335	44.942	0.359
16	16	24	32	38	42	46	50	54	58	60	60	1.426	46.209	0.332
17	18	26	34	40	44	47	51	55	59	60	60	1.517	47.466	0.306
Girls														
9	5	8	12	15	18	21	25	29	35	43	51	0.394	21.258	0.642
10	7	10	14	18	22	26	30	34	40	50	58	0.485	25.666	0.605
11	8	11	17	21	25	29	34	39	45	54	60	0.571	29.444	0.569
12	9	13	19	24	28	32	37	42	48	57	60	0.646	32.123	0.534
13	10	15	21	26	30	34	39	44	50	59	60	0.705	34.408	0.504
14	11	15	22	27	31	35	40	45	50	59	60	0.741	35.334	0.482
15	11	16	22	27	31	35	40	44	50	58	60	0.757	35.327	0.464
16	12	17	23	28	32	36	40	44	50	57	60	0.761	35.690	0.447
17	13	18	24	28	32	36	40	45	50	58	60	0.761	36.333	0.431

Percentile data were calculated from 8 837 sit-up test performances collected between 1985 and 1999.

L, skew; M, median; P, percentile; S, coefficient of variation.

Table 9 Hand-grip strength (kg) centile values and LMS summary statistics by sex and age in 9- to 15-year-old Australians (taken as the mean of both hands)

Age (year)	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅	L	M	S
Boys														
9	11.5	12.5	13.8	14.8	15.6	16.4	17.2	18.1	19.2	20.8	22.1	0.600	16.415	0.197
10	13.1	14.3	15.9	17.0	18.0	19.0	19.9	21.0	22.2	23.9	25.4	0.728	18.967	0.198
11	14.5	15.9	17.7	19.0	20.1	21.2	22.3	23.5	24.9	26.8	28.5	0.764	21.217	0.200
12	15.4	17.0	18.9	20.3	21.5	22.7	23.8	25.1	26.6	28.7	30.5	0.747	22.655	0.203
13	17.5	19.3	21.5	23.1	24.5	25.8	27.2	28.6	30.4	32.8	34.9	0.738	25.819	0.205
14	20.8	22.9	25.5	27.4	29.1	30.7	32.4	34.1	36.2	39.1	41.6	0.742	30.731	0.207
15	24.6	27.1	30.3	32.6	34.6	36.5	38.4	40.5	43.0	46.5	49.5	0.752	36.517	0.207
Girls														
9	9.8	10.8	12.0	12.9	13.7	14.4	15.1	16.0	17.0	18.4	19.5	0.639	14.396	0.205
10	11.4	12.6	14.1	15.2	16.2	17.1	18.0	19.0	20.1	21.8	23.1	0.842	17.072	0.210
11	12.5	13.9	15.5	16.8	17.8	18.8	19.8	20.9	22.1	23.9	25.3	0.932	18.816	0.208
12	14.4	16.0	17.8	19.1	20.3	21.4	22.5	23.6	25.0	26.9	28.5	0.922	21.374	0.200
13	16.4	18.0	19.9	21.3	22.5	23.6	24.8	26.0	27.4	29.5	31.1	0.880	23.641	0.190
14	18.2	19.7	21.6	23.0	24.3	25.4	26.5	27.8	29.2	31.3	33.0	0.828	25.390	0.178
15	19.8	21.3	23.2	24.6	25.8	26.9	28.0	29.2	30.7	32.7	34.4	0.770	26.881	0.165

Percentile data were calculated from the 3 707 hand-grip strength performances collected between 1985 and 1999.

L, skew; M, median; P, percentile; S, coefficient of variation.

were generally small, with Australian children scoring slightly higher on hand-grip strength (mean $\pm$ 95% CI: 0.20 ± 0.03 SDs) and 50 m sprint tests (0.24 ± 0.02 SDs), and slightly lower on sit-and-reach (-0.36 ± 0.02 SDs), standing broad jump (-0.25 ± 0.02 SDs) and 20 m shuttle run tests (-0.49 ± 0.01 SDs).

DISCUSSION

This study provides the most up-to-date sex- and age-specific normative centile values for 9–17-year-old Australians across a range of health-related fitness tests, which can be used as benchmark values for health and fitness screening

and surveillance of children. These data complement a growing literature reporting growth percentiles across a range of different health measures, such as body mass index,³² waist girth³³ and blood pressure,²⁸ and a range of other health-related fitness measures.^{29–30} It also quantifies the magnitude and direction of sex- and age-related differences in children's health-related fitness and shows that boys consistently scored higher than girls on fitness tests (except on the sit-and-reach test of flexibility) and that boys experience larger age-related changes in fitness. The developmental patterns of children's fitness have been well studied and extensively reviewed (eg, for cardiovascular fitness, refer to Armstrong

Original Article

Table 10 Sit-and-reach (cm) centile values and LMS summary statistics by sex and age in 9- to 17-year-old Australians.

Age (y)	P ₅	P ₁₀	P ₂₀	P ₃₀	P ₄₀	P ₅₀	P ₆₀	P ₇₀	P ₈₀	P ₉₀	P ₉₅	L	M	S
<i>boys</i>														
9	10.4	12.9	15.7	17.7	19.4	20.9	22.4	23.9	25.8	28.2	30.3	1.211	20.877	0.285
10	10.0	12.5	15.3	17.3	19.0	20.5	22.1	23.7	25.5	28.0	30.1	1.190	20.537	0.294
11	9.6	12.1	15.0	17.0	18.7	20.3	21.9	23.5	25.4	28.0	30.1	1.167	20.313	0.305
12	9.3	11.8	14.8	16.9	18.7	20.3	21.9	23.6	25.6	28.3	30.5	1.133	20.292	0.315
13	9.4	12.0	15.1	17.2	19.1	20.8	22.5	24.3	26.4	29.2	31.6	1.091	20.785	0.322
14	9.8	12.5	15.7	18.0	20.0	21.8	23.6	25.5	27.8	30.9	33.4	1.054	21.804	0.328
15	10.4	13.2	16.6	19.1	21.2	23.1	25.1	27.1	29.5	32.9	35.7	1.017	23.112	0.332
16	11.1	14.0	17.6	20.1	22.3	24.4	26.5	28.7	31.3	34.9	37.8	0.984	24.392	0.334
17	11.7	14.8	18.5	21.2	23.5	25.7	27.9	30.2	33.0	36.8	40.0	0.953	25.686	0.335
<i>girls</i>														
9	13.0	15.8	18.9	21.1	22.9	24.6	26.2	28.0	29.9	32.6	34.8	1.285	24.614	0.264
10	13.0	15.7	18.8	20.9	22.7	24.4	26.0	27.7	29.7	32.4	34.6	1.259	24.402	0.265
11	13.2	15.9	19.0	21.2	23.0	24.7	26.4	28.1	30.1	32.8	35.0	1.235	24.705	0.265
12	14.0	16.7	19.9	22.2	24.1	25.8	27.5	29.3	31.3	34.2	36.4	1.230	25.790	0.262
13	15.3	18.2	21.6	24.0	25.9	27.7	29.5	31.4	33.6	36.5	38.9	1.250	27.740	0.256
14	16.5	19.5	23.1	25.5	27.6	29.4	31.3	33.2	35.4	38.4	40.9	1.293	29.440	0.248
15	17.0	20.1	23.7	26.1	28.1	30.0	31.8	33.7	35.9	38.8	41.2	1.350	29.997	0.241
16	17.0	20.0	23.5	25.9	27.9	29.6	31.4	33.2	35.3	38.1	40.3	1.412	29.647	0.235
17	16.8	19.8	23.2	25.5	27.4	29.1	30.7	32.5	34.4	37.1	39.2	1.472	29.074	0.229

Note, percentile data were calculated from 9,124 sit-and-reach performances collected between 1985 and 2000; L = skew; M = median; S = coefficient of variation. Note, a score of "20 cm" corresponds to the participant reaching their toes.

et al,³⁴ Krahenbuhl *et al*³⁵ and Rowland³⁶; for muscular fitness, refer to Blimkie and Sale,³⁷ Froberg and Lammert³⁸ and De Ste Croix³⁹). Although the underlying causes of sex- and age-related differences are clear for some fitness test performances, such as those for muscular strength, power and speed, which are largely explained by physical differences (eg, differences in muscle mass or height), they are less clear for others, such as for cardiovascular endurance, which may be explained by physiological differences (eg, differences in mechanical efficiency and/or fractional utilisation).^{15 36} It is, nonetheless, beyond the scope of this article to discuss the causes that underscore the sex- and age-related changes in fitness test performance.

International comparisons

Although several studies have previously compared the health-related fitness of Australian children with their sex- and age-matched international peers,^{20 40} comparisons have only been made for cardiovascular fitness. Figure 6 compares the 20-m shuttle run, 50 m sprint, standing broad jump, hand-grip strength and sit-and-reach performance of 9–17-year-old Australians with 1 894 971 test results from sex-, age- and test-matched international children from 48 countries who have been measured using the same test protocols as those referenced in table 1 and described in Appendix 1. Figure 6 also shows typically small differences in health-related fitness between Australian and international children. Furthermore, the sex- and age-related differences in fitness of Australian children are strikingly similar to those observed in international children. Given that the differences are generally small, the normative centile data presented in this study could be used as approximate benchmark values for health-related fitness of international children.

Fitness thresholds for cardiometabolic risk

Fitness is widely recognised as a powerful marker of current and future cardiovascular, skeletal and mental health.

Unfortunately, there are no universally accepted recommendations for health-related levels of fitness. In recent years however, sex- and age-specific threshold values for cardiovascular fitness (operationalised as mass-specific peak oxygen uptake in ml/kg/min) have been established for European and US children using linked cardiometabolic risk-based values from receiver operator characteristic curve analyses.^{41–44} To estimate the prevalence of Australian children with 'healthy' cardiovascular fitness (ie, those above the thresholds), 'international' sex- and age-specific thresholds for 9–17-year-old children were estimated by determining best-fitting polynomial regression model (quadratic or cubic) relating age (predictor variable) to previously reported threshold values (response variable) in Adegboye *et al*,⁴¹ Lobelo *et al*,⁴² Ruiz *et al*⁴³ and Welk *et al*.⁴⁴ Separate models were generated for boys and girls. Peak oxygen uptake values in Australian children were estimated using 1.6 km run and 20 m shuttle run data and the Cureton *et al*⁴⁵ and Léger *et al*⁴⁶ regression equations, respectively.

Using these thresholds, about 71% of Australian boys (median $\pm 95\%$ CI: 71% $\pm 8\%$) and 77% of Australian girls (median $\pm 95\%$ CI: 77% $\pm 10\%$) apparently have 'healthy' cardiovascular fitness. Although in light of recent secular declines in cardiovascular fitness,^{20 22 23 25} and with a median testing year of 1993 in this study's cardiovascular fitness dataset, it is likely that these prevalence rates somewhat overestimate those of today. These prevalence rates are better than (for girls), or similar to (for boys), those observed in European (61% of boys and 57% of girls)²⁹ and US (71% of boys and 69% of girls)⁴² children. Geographical differences in prevalence rates may reflect differences in (a) threshold levels, (b) the year(s) of testing, (c) sampling methodology, (d) test methodology and (e) the way in which peak oxygen uptake was measured or estimated.⁴⁷

Ultimately, it is important to remember that the normative data presented in this study show how well Australian children perform on health-related fitness tests relative to their sex- and age-matched peers. For example, using a percentile

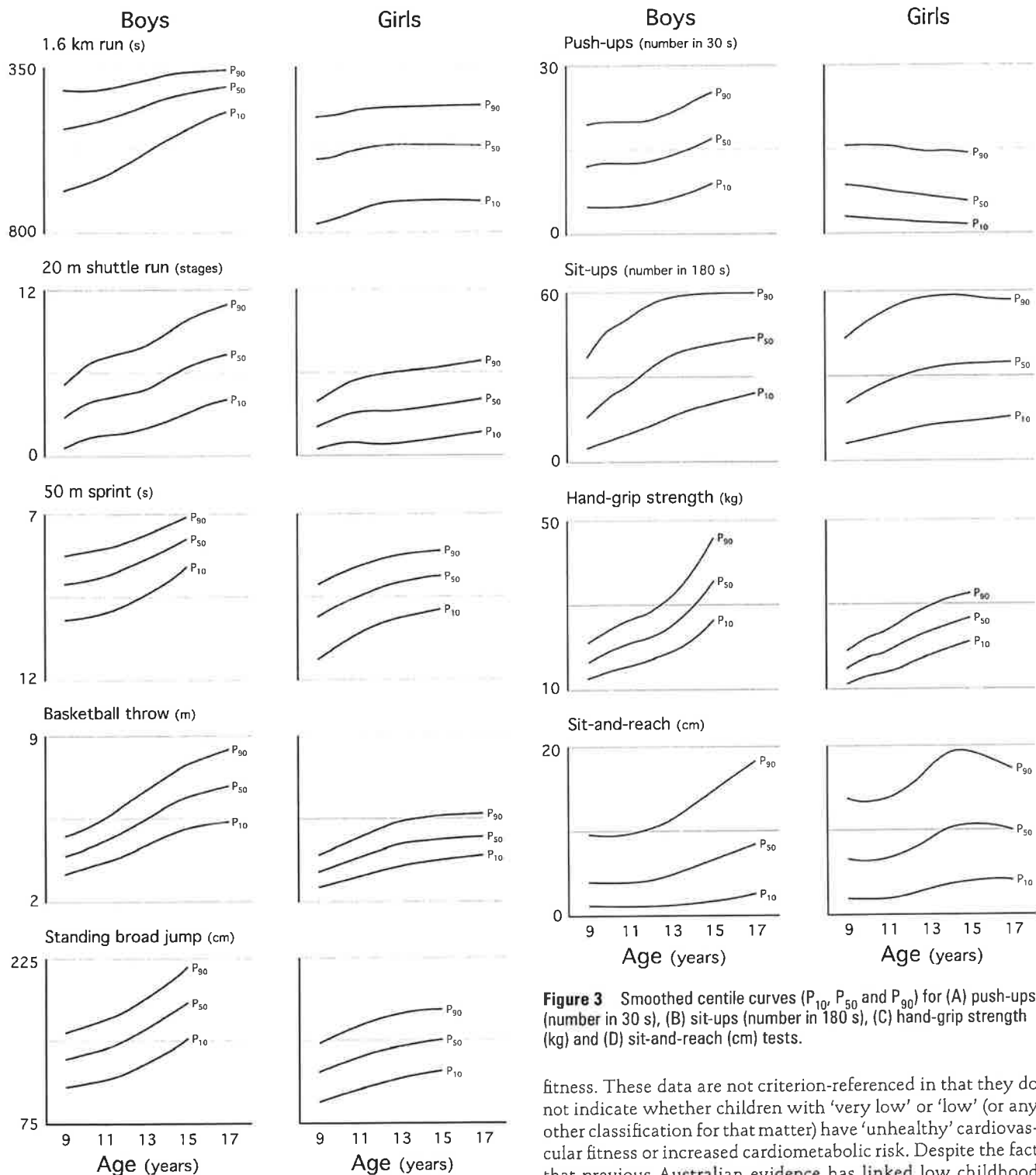


Figure 2 Smoothed centile curves (P₁₀, P₅₀ and P₉₀) for (A) 1.6 km run (s), (B) 20 m shuttle run (number of completed stages), (C) 50 m sprint (s), (D) basketball throw (m) and (E) standing broad jump (cm).

classification, children with fitness in the bottom 20% can be classified as having 'very low' fitness; those between the 20th and 40th percentiles as having 'low' fitness; those between the 40th and 60th percentiles as having 'average' fitness; those between the 60th and 80th percentiles as having 'high' fitness; and those above the 80th percentile as having 'very high'

Figure 3 Smoothed centile curves (P₁₀, P₅₀ and P₉₀) for (A) push-ups (number in 30 s), (B) sit-ups (number in 180 s), (C) hand-grip strength (kg) and (D) sit-and-reach (cm) tests.

fitness. These data are not criterion-referenced in that they do not indicate whether children with 'very low' or 'low' (or any other classification for that matter) have 'unhealthy' cardiovascular fitness or increased cardiometabolic risk. Despite the fact that previous Australian evidence has linked low childhood cardiovascular fitness with increased cardiometabolic risk in adulthood,⁴⁸ future Australian studies are required to examine whether childhood thresholds for cardiovascular fitness (or other health-related fitness components) are significantly associated with clustered cardiometabolic risk (or other health outcomes, such as mental or skeletal health outcomes).

Strengths and limitations

Despite the fact that the last national fitness survey of Australian children was in 1985, this study provides the most

Original Article

Effect size

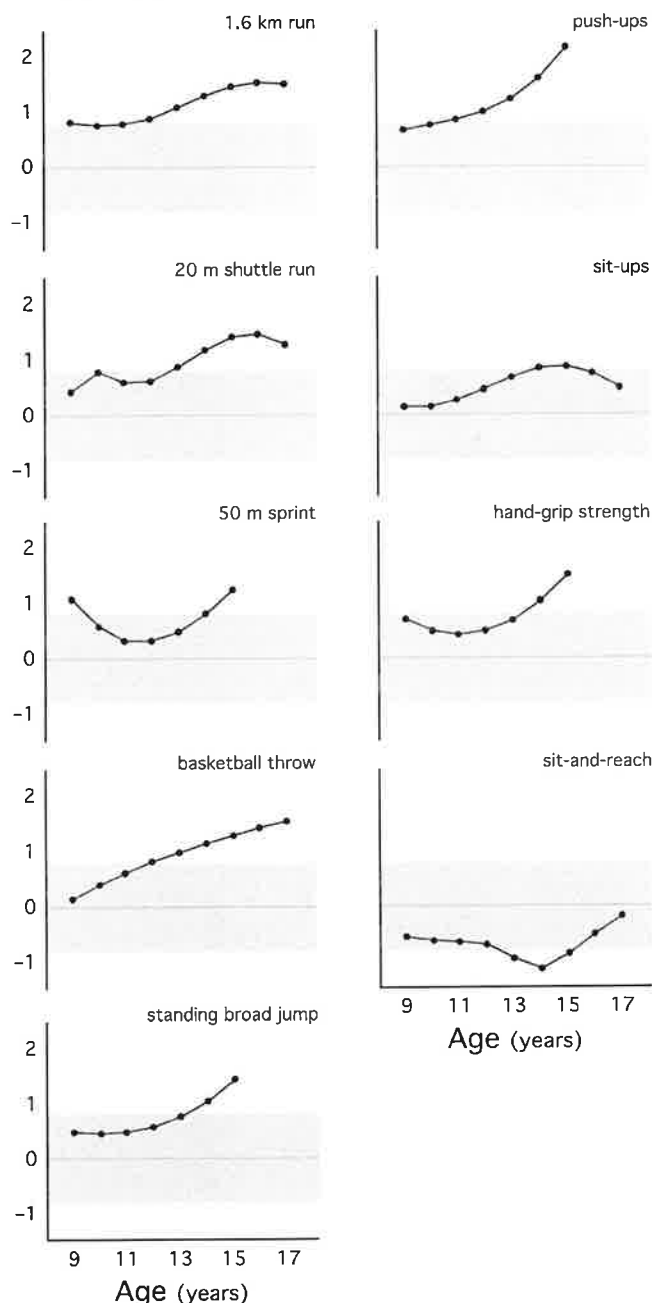


Figure 4 Sex-related differences in mean fitness expressed as effect sizes. Data are shown for 9–17-year-old children tested on the (A) 1.6 km run, (B) 20 m shuttle run, (C) 50 m sprint, (D) basketball throw, (E) standing broad jump, (F) push-ups, (G) sit-ups, (H) hand-grip strength and (I) sit-and-reach tests. The limits of the grey zone represent effects sizes of 0.8 and –0.8, beyond which large differences are observed.

up-to-date normative dataset for nine widely administered health-related fitness tests, using cumulated data from 85347 Australian children aged 9–17 years collected between 1985 and 2009. This study used a strict set of inclusion and exclusion criteria and rigorous initial data analysis procedures to systematically control for any factors (eg, differences in test methodology) that might have biased the normative values or the estimates of the sex- and age-related differences. It used

Effect size (age 15 years = 0)

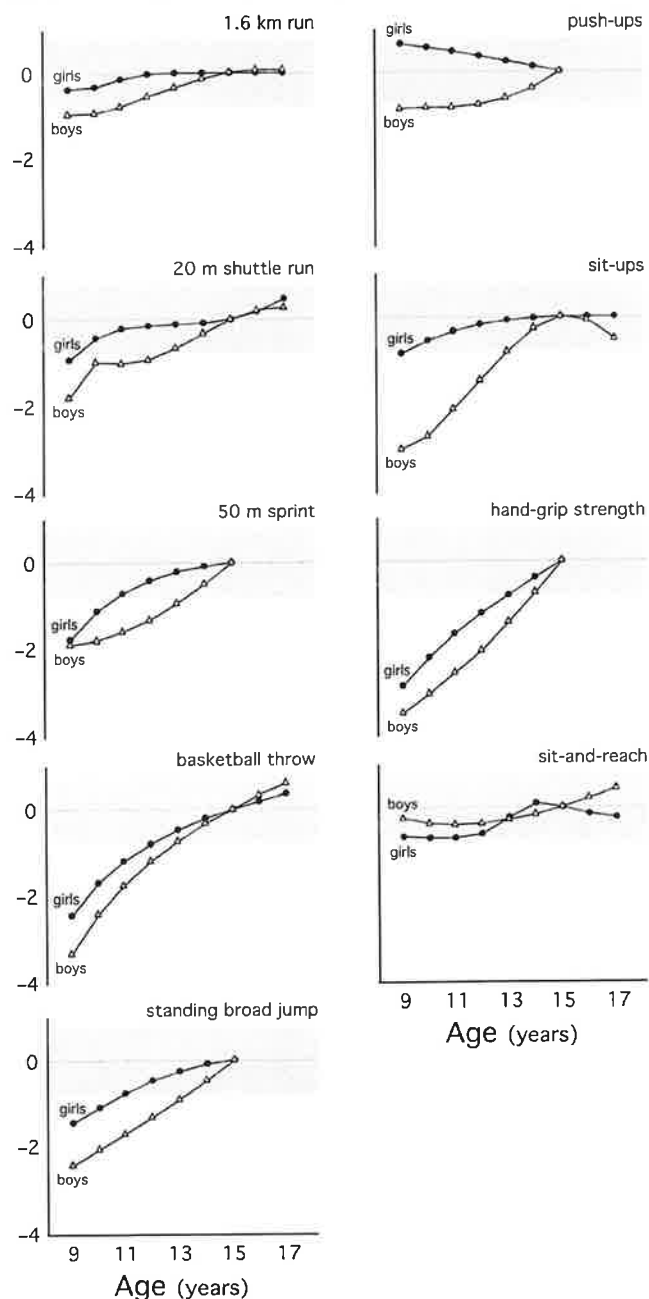


Figure 5 Age-related changes in mean fitness expressed as effect sizes standardised to an effect size of age 15 years = 0. Data are shown for 9–17-year-old boys (triangles) and girls (circles) separately tested on the (A) 1.6 km run, (B) 20 m shuttle run, (C) 50 m sprint, (D) basketball throw, (E) standing broad jump, (F) push-ups, (G) sit-ups, (H) hand-grip strength and (I) sit-and-reach tests. The limits of the grey zone represent effects sizes of 0.8 and –0.8, beyond which large differences are observed.

a novel pseudo-data method to allow both descriptive and raw data to be merged before using the LMS method to create sex- and age-specific smoothed percentiles. It also quantified sex- and age-related differences as standardised effect sizes, allowing for comparison between sexes, among different ages, and with sex, age and test-matched international children.

Effect size

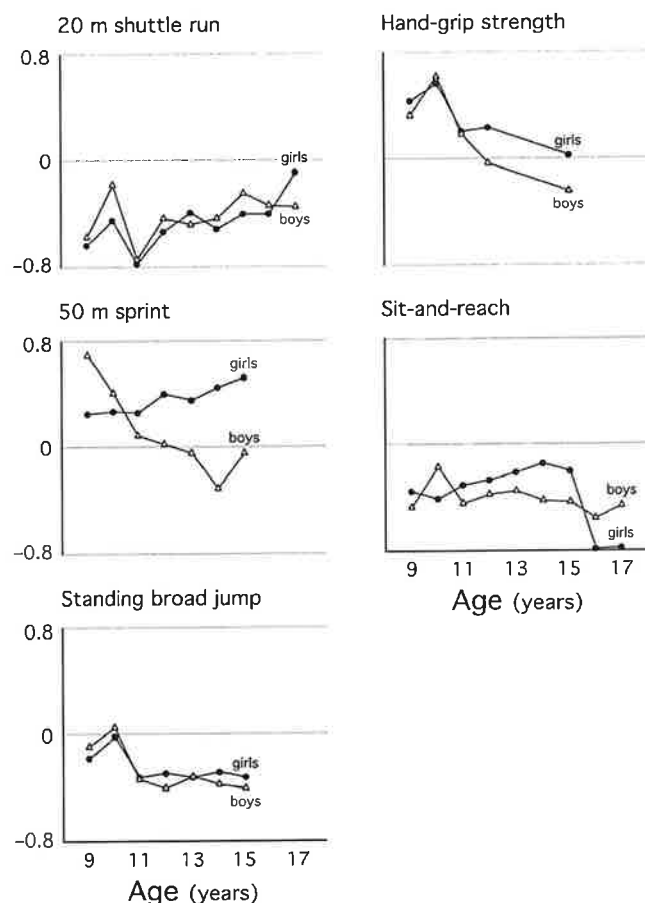


Figure 6 Sex- and age-specific effect sizes for (A) 20 m shuttle run, (B) 50 m sprint, (C) hand-grip strength, (D) sit-and-reach and (E) standing broad jump for 9–17-year-old Australian boys (triangles) and girls (circles) relative to their international peers. Positive effects indicate higher fitness scores for Australian children and negative effects indicate lower fitness scores. Comparative data represent $n=284\ 508$ 20 m shuttle run performances,^{28–30} $n=1\ 216\ 452$ 50 m sprint performances,¹⁸ $n=126\ 361$ hand-grip strength performances,²⁹ $n=102\ 664$ sit-and-reach performances,²⁹ and $n=164\ 986$ standing broad jump performances²⁹ of 9–17-year-old children from 48 international countries.

However, this study is not without limitations. Only one of the 15 included studies was based on a nationally representative sample, which obviously raises the issue of representativeness. Most of the included studies used similar sampling frames (table 1). Schools with a greater interest in sport and fitness may have been more willing to participate, and because participation at the individual level was voluntary, it is possible that children with low fitness levels chose not to participate. This might have resulted in fitness test performances unrepresentative of the population, but it should not have affected the sex- and age-related differences. Fitness data were also collected at different times during the 1985–2009 period, and given convincing evidence of recent temporal declines in some (but not all) components of Australian children's fitness,^{23 49} it is possible that the normative data presented in this study represent a better 'health-related picture' than what would be observed today. A temporal analysis of the data accumulated in this study suggests that these normative data would probably

overestimate the fitness of Australian children in 2009 by an average of 0.3 SDs or 13 percentile points, assuming of course that the observed temporal changes remained consistent across the entire 1985–2009 period. Nonetheless, these data represent the best available and most up-to-date health-related fitness data on Australian children. It must also be remembered that despite being simple, cheap, easy, reliable, reasonably valid and widely used alternatives of laboratory-based criterion measures, field tests are affected by factors other than underlying construct fitness. For example, validity data for field tests of cardiovascular fitness suggest that (at best) only 50–60% of the variance in field test performance is explained by the variance in underlying peak oxygen uptake, indicating that other physiological, physical, biomechanical, psychosocial and environmental factors also play a part.¹⁵ In addition, although criterion-related validity has not been established for all of the included tests, face validity is generally accepted.¹⁷ Most of the included tests are also considered to demonstrate good reliability, although tests requiring a reasonable degree of subjective judgement (eg, the subjective scoring of a properly performed sit-up or push-up) typically demonstrate poorer reliability.¹⁴

CONCLUSION

Physical fitness is considered to be an excellent marker of current and future health. In anticipation of a follow-up national fitness survey, this study provides the most up-to-date and most comprehensive set of sex- and age-specific normative centile values of health-related fitness of Australian children, which can be used as benchmark values for health and fitness screening and surveillance systems. These normative centile values will facilitate the identification of children with low fitness to set appropriate fitness goals, monitor individual changes in fitness and promote positive health behaviours. They will also facilitate the identification of children who possess specific fitness characteristics that may be considered important for sporting success, in the hope of recruiting the high achievers into elite sporting development programs.

Correction notice This article has been corrected since it was published Online First. The authors have noticed that the normative data in Table 10 are incorrect. The correct table has been inserted.

Acknowledgements The authors thank the authors of the included studies for generously clarifying details of their studies and/or for providing raw data. The University of South Australia Divisional Development Research Scheme supported this study.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- Ortega FB, Ruiz JR, Castillo MJ, et al. Physical fitness in childhood and adolescence: a powerful marker of health. *Int J Obes (Lond)* 2008;**32**:1–11.
- Ruiz JR, Castro-Piñero J, Artero EG, et al. Predictive validity of health-related fitness in youth: a systematic review. *Br J Sports Med* 2009;**43**:909–23.
- Vicente-Rodriguez G, Ara I, Perez-Gomez J, et al. High femoral bone mineral density accretion in prepubertal soccer players. *Med Sci Sports Exerc* 2004;**36**:1789–95.
- Kodama S, Saito K, Tanaka S, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. *JAMA* 2009;**301**:2024–35.
- Lee CD, Blair SN. Cardiorespiratory fitness and stroke mortality in men. *Med Sci Sports Exerc* 2002;**34**:592–5.
- Dishman RK, Washburn RA, Heath GW. *Physical Activity Epidemiology*. Champaign, IL: Human Kinetics 2004:358–9.
- Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev* 2009;**2**:CD007146.
- Katzmarzyk PT, Church TS, Janssen I, et al. Metabolic syndrome, obesity, and mortality: impact of cardiorespiratory fitness. *Diabetes Care* 2005;**28**:391–7.

Original Article

9. **Katzmarzyk PT**, Church TS, Blair SN. Cardiorespiratory fitness attenuates the effects of the metabolic syndrome on all-cause and cardiovascular disease mortality in men. *Arch Intern Med* 2004;**164**:1092–7.
10. **Andersen LB**, Hasselstrom H, Grønfeldt V, *et al*. The relationship between physical fitness and clustered risk, and tracking of clustered risk from adolescence to young adulthood: eight years follow-up in the Danish Youth and Sport Study. *Int J Behav Nutr Phys Act* 2004;**1**:6.
11. **Cleland VJ**, Ball K, Magnussen C, *et al*. Socioeconomic position and the tracking of physical activity and cardiorespiratory fitness from childhood to adulthood. *Am J Epidemiol* 2009;**170**:1069–77.
12. **Twisk JW**, Kemper HC, van Mechelen W. Tracking of activity and fitness and the relationship with cardiovascular disease risk factors. *Med Sci Sports Exerc* 2000;**32**:1455–61.
13. **Kristensen PL**, Wedderkopp N, Møller NC, *et al*. Tracking and prevalence of cardiovascular disease risk factors across socio-economic classes: a longitudinal substudy of the European Youth Heart Study. *BMC Public Health* 2006;**6**:20.
14. **Artero EG**, España-Romero V, Castro-Piñero J, *et al*. Reliability of field-based fitness tests in youth. *Int J Sports Med* 2011;**32**:159–69.
15. **Tomkinson GR**, Olds TS. Field tests of fitness. In: Armstrong N, Van Mechelen W, eds. *Pediatric Exercise Science and Medicine*. Oxford: Oxford University Press 2008:109–28.
16. **Castro-Piñero J**, Artero EG, España-Romero V, *et al*. Criterion-related validity of field-based fitness tests in youth: a systematic review. *Br J Sports Med* 2010;**44**:934–43.
17. **Ruiz JR**, Castro-Piñero J, España-Romero V, *et al*. Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. *Br J Sports Med* 2011;**45**:518–24.
18. Council of Europe. *Eurofit: Handbook for the Eurofit Tests Of Physical Fitness*. Rome: Council of Europe 1988.
19. The Cooper Institute. *FITNESSGRAM/ACTIVITYGRAM Test Administration Manual*. Fourth edition. Champaign, IL: Human Kinetics 2007.
20. **Olds TS**, Tomkinson GR. The aerobic fitness of 9–15 year old South Australian children: norms, trends and international comparisons. *ACHPER Healthy Lifestyles Journal* 2003;**50**:25–30.
21. **Pyke JE**. *Australian Health And Fitness Survey 1985: The Fitness, Health And Physical Performance Of Australian School Students Aged 7–15 Years*. Parkside, SA: The Australian Council for Health, Physical Education and Recreation 1987.
22. **Tomkinson GR**, Olds TS. Secular changes in pediatric aerobic fitness test performance: the global picture. *Med Sport Sci* 2007;**50**:46–66.
23. **Tomkinson GR**, Olds TS. Secular changes in aerobic fitness test performance of Australasian children and adolescents. *Med Sport Sci* 2007;**50**:168–82.
24. **Tomkinson GR**. Global changes in anaerobic fitness test performance of children and adolescents (1958–2003). *Scand J Med Sci Sports* 2007;**17**:497–507.
25. **Tomkinson GR**, Léger LA, Olds TS, *et al*. Secular trends in the performance of children and adolescents (1980–2000): an analysis of 55 studies of the 20m shuttle run test in 11 countries. *Sports Med* 2003;**33**:285–300.
26. **Cole TJ**, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992;**11**:1305–19.
27. **Pan H**, Cole T. *User's Guide to Lms Chart Maker*. Council MR, ed. 1997–2005 Medical Research Council. UK: Medical Research Council 2005.
28. **Jackson LV**, Thalange NK, Cole TJ. Blood pressure centiles for Great Britain. *Arch Dis Child* 2007;**92**:298–303.
29. **Ortega FB**, Artero EG, Ruiz JR, *et al*. Physical fitness levels among European adolescents: the HELENA study. *Br J Sports Med* 2011;**45**:20–9.
30. **Castro-Piñero J**, González-Montesinos JL, Keating XD, *et al*. Percentile values for running sprint field tests in children ages 6–17 years: influence of weight status. *Res Q Exerc Sport* 2010;**81**:143–51.
31. **Cohen J**. *Statistical Power Analysis For The Behavioral Sciences*. Second edition. New Jersey: Lawrence Erlbaum 1988.
32. **Cole TJ**, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990. *Arch Dis Child* 1995;**73**:25–9.
33. **Eisenmann JC**. Waist circumference percentiles for 7- to 15-year-old Australian children. *Acta Paediatr* 2005;**94**:1182–5.
34. **Armstrong N**, McManus AM, Welsman JR. Aerobic fitness. In: Armstrong N, Van Mechelen W, eds. *Paediatric Exercise Science and Medicine*. Oxford: Oxford University Press 2008:269–82.
35. **Krahenbuhl GS**, Skinner JS, Kohrt WM. Developmental aspects of maximal aerobic power in children. *Exerc Sport Sci Rev* 1985;**13**:503–38.
36. **Rowland TW**. Evolution of maximal oxygen uptake in children. *Med Sport Sci* 2007;**50**:200–9.
37. **Blimkie CJR**, Sale DG. Strength development and trainability during childhood. In: Van Praagh E, eds. *Pediatric Anaerobic Performance*. Champaign, IL: Human Kinetics 1998:193–224.
38. **Froberg K**, Lammert O. Development of muscle strength during childhood. In: Bar-Or O, ed. *The Child and Adolescent Athlete*. Oxford: Blackwell Science 1996:42–53.
39. **De Ste Croix MDA**. Muscle strength. In: Armstrong N, Van Mechelen W, eds. *Paediatric Exercise Science and Medicine*. Second edition. Oxford: Oxford University Press 2008:199–212.
40. **Olds T**, Tomkinson G, Léger L, *et al*. Worldwide variation in the performance of children and adolescents: an analysis of 109 studies of the 20-m shuttle run test in 37 countries. *J Sports Sci* 2006;**24**:1025–38.
41. **Adegboye AR**, Anderssen SA, Froberg K, *et al*. Recommended aerobic fitness level for metabolic health in children and adolescents: a study of diagnostic accuracy. *Br J Sports Med* 2011;**45**:722–8.
42. **Lobelo F**, Pate RR, Dowda M, *et al*. Validity of cardiorespiratory fitness criterion-referenced standards for adolescents. *Med Sci Sports Exerc* 2009;**41**:1222–9.
43. **Ruiz JR**, Ortega FB, Rizzo NS, *et al*. High cardiovascular fitness is associated with low metabolic risk score in children: the European Youth Heart Study. *Pediatr Res* 2007;**61**:350–5.
44. **Welk GJ**, Laurson KR, Eisenmann JC, *et al*. Development of youth aerobic capacity standards using receiver operator characteristic curves. *Am J Prev Med* 2011;**41**:S111–6.
45. **Cureton KJ**, Sloniger MA, O'Bannon JP, *et al*. A generalized equation for prediction of VO₂peak from 1-mile run/walk performance. *Med Sci Sports Exerc* 1995;**27**:445–51.
46. **Léger L**, Lambert J, Goulet A, *et al*. [Aerobic capacity of 6 to 17-year-old Quebecois—20 meter shuttle run test with 1 minute stages]. *Can J Appl Sport Sci* 1984;**9**:64–9.
47. **Tomkinson G**. Aerobic fitness thresholds for cardio metabolic health in children and adolescents. *Br J Sports Med* 2011;**45**:686–7.
48. **Dwyer T**, Magnussen CG, Schmidt MD, *et al*. Decline in physical fitness from childhood to adulthood associated with increased obesity and insulin resistance in adults. *Diabetes Care* 2009;**32**:683–7.
49. **Tomkinson GR**, Hamlin MJ, Olds TS. Secular changes in anaerobic test performance in Australasian children and adolescents. *Pediatr Exerc Sci* 2006;**18**:314–28.
50. Australian Council for Health, Physical Education and Recreation. *Australian Fitness Education Award: User's Manual And Curriculum Ideas*. Adelaide, SA: Australian Council for Health, Physical Education and Recreation 1996.
51. **Barnett LM**, Van Beurden E, Morgan PJ, *et al*. Does childhood motor skill proficiency predict adolescent fitness? *Med Sci Sports Exerc* 2008;**40**:2137–44.
52. **Birchall J**. Health related fitness testing. In: Went S, ed. *A Healthy Start: Holistic Approaches to Health Promotion in School Communities*. Second edition. Melbourne: Monash University 1992:251–65.
53. **Booth M**, Macaskill P, McLellan L, *et al*. NSW Schools Fitness and Physical Activity Survey 1997. Sydney: NSW Department of Education and Training, 1997.
54. **Booth M**, Okely AD, Denney-Wilson E, *et al*. NSW Schools Physical Activity and Nutrition Survey (SPANS) 2004: Full Report. Sydney: NSW Department of Health, 2006.
55. **Burke V**, Beilin LJ, Durkin K, *et al*. Television, computer use, physical activity, diet and fatness in Australian adolescents. *Int J Pediatr Obes* 2006;**1**:248–55.
56. **Cooley D**, McNaughton L. Aerobic fitness of Tasmanian secondary school children using the 20-m shuttle run test. *Percept Mot Skills* 1999;**88**:188–98.
57. **Dollman J**, Olds T, Norton K, *et al*. The evolution of fitness and fatness in 10–11-year-old Australian schoolchildren: changes in distributional characteristics between 1985 and 1997. *Pediatr Exerc Sci* 1999;**11**:106–21.
58. **Hands B**. *Fitness and motor skill levels of Western Australian Primary School Children*. Perth, WA: University of Western Australia 2000.
59. **McNaughton L**, Morgan R, Smith P, *et al*. An investigation into the fitness levels of Tasmanian primary schoolchildren. *ACHPER Healthy Lifestyles Journal* 1996;**43**:4–10.
60. **Vandongen R**, Jenner DA, Thompson C, *et al*. A controlled evaluation of a fitness and nutrition intervention program on cardiovascular health in 10- to 12-year-old children. *Prev Med* 1995;**24**:9–22.



Normative health-related fitness values for children: analysis of 85347 test results on 9–17-year-old Australians since 1985

Mark J Catley and Grant R Tomkinson

Br J Sports Med published online October 21, 2011

Updated information and services can be found at:
<http://bjsm.bmj.com/content/early/2012/03/26/bjsports-2011-090218>

These include:

Supplementary Material

Supplementary material can be found at:
<http://bjsm.bmj.com/content/suppl/2011/10/21/bjsports-2011-090218.DC1.html>

References

This article cites 43 articles, 11 of which you can access for free at:
<http://bjsm.bmj.com/content/early/2012/03/26/bjsports-2011-090218#BIBL>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:
<http://group.bmj.com/subscribe/>

Shanlee M. Davis*, Jill L. Kaar, Brandy M. Ringham, Christine W. Hockett, Deborah H. Glueck and Dana Dabelea

Sex differences in infant body composition emerge in the first 5 months of life

<https://doi.org/10.1515/jpem-2019-0243>

Received June 4, 2019; accepted July 26, 2019; previously published online September 4, 2019

Abstract

Background: Sex differences in body composition are appreciated throughout the lifespan with probable contributions from sex steroids: testosterone and estrogen. The purpose of this longitudinal observational study was to determine if sex differences in body composition emerge during the first months of life in healthy infants, corresponding to the age at which male infants produce endogenous testosterone.

Methods: Linear growth and body composition parameters using air displacement plethysmography were obtained from 602 healthy infants after birth and again at 5 months of age. Rate of change in body composition parameters were compared between sexes.

Results: Sex differences in length, total mass, fat free mass (FFM), and percent fat mass (%FM) were present both at birth and at 5 months ($p < 0.001$ for all), with males having greater total mass and FFM but lower %FM. Gain in %FM over the first 5 months was significantly lower in males ($p = 0.0004$). This difference was secondary to a gain of 17 g/week more in FFM in males compared to females.

Conclusions: Sex differences in body composition emerge in the first months of life, with lower adiposity accumulation in males. Endogenous testosterone production in males ~1–4 months of age may account for findings and

may have lifelong implications for sex differences in body composition.

Keywords: adipose tissue; body composition; body fat percentage; critical window; mini-puberty; PEA POD.

Background

Sex differences in body composition are recognized throughout the lifespan, with females having significantly lower lean mass and greater percent fat mass (%FM) than males. Sex steroids (testosterone and estrogen) are assumed to be largely responsible for these differences; however, sex differences are observed in childhood well before puberty occurs [1]. While some studies report that sex differences are present at birth [2–4] others have found no differences between males and females [5, 6]. By 1 year of age, females have greater %FM and this difference seems to persist until puberty at which time it becomes further exaggerated. Although sex differences in body composition are known to exist, the underlying etiology and timing of emergence are surprisingly understudied.

The mini-puberty period of infancy is a biologically plausible time during which sex differences in body composition may emerge. The mini-puberty period refers to temporary activation of the hypothalamic-pituitary-gonadal axis resulting in near-adult concentrations of testosterone in boys from around 1–4 months of age [7]. Testosterone is an anabolic steroid that acts to increase protein synthesis and lean mass while decreasing fat mass (FM). Although the existence of the mini-puberty period is well known by pediatric endocrinologists, the short and long-term consequences of this temporary testosterone exposure are just beginning to be unveiled. It was recently shown that there are sex differences in linear growth velocity during the first 6 months of life associated with testosterone concentrations [8]. This difference in infancy was calculated to be substantial enough to account for 15% of the final height differences between men and women [9]. Animal models have found that blocking the post-natal testosterone surge in male mice results in greater adiposity lifelong [10]. The effects of sex steroid exposure during

*Corresponding author: Shanlee M. Davis, MD, MS, University of Colorado, Department of Pediatrics, Denver, CO 80203, USA; and Children's Hospital Colorado, 13123 East 16th Ave B265, Aurora, CO 80045, USA, Phone: +720-777-6128, Fax: +720-777-7301, E-mail: Shanlee.davis@childrenscolorado.org.
<https://orcid.org/0000-0002-0304-9550>

Jill L. Kaar and Deborah H. Glueck: University of Colorado, Department of Pediatrics, Aurora, CO, USA

Brandy M. Ringham: University of Colorado, Lifecourse Epidemiology of Adiposity and Diabetes Center, Aurora, CO, USA

Christine W. Hockett: University of Colorado, Department of Epidemiology, Aurora, CO, USA

Dana Dabelea: University of Colorado, Lifecourse Epidemiology of Adiposity and Diabetes Center, Aurora, CO, USA; and University of Colorado, Department of Epidemiology, Aurora, CO, USA

the mini-puberty period on body composition in humans have not been studied.

The objective of this secondary data analysis was to determine if sex differences in body composition emerge during the first months of life at a sensitive time period at which sex steroid exposure is known to be substantially different between the sexes. We hypothesize that differences in body composition emerge during the first 5 months of life, with males developing more lean mass and lower %FM in a pattern consistent with testosterone exposure.

Materials and methods

Between 2009 and 2014, the Healthy Start Study enrolled 1410 mother-offspring dyads before 24 weeks' gestation from obstetrics clinics at the University of Colorado Hospital. Detailed assessment of the study participants and methods have been previously published [11–14]. In brief, sociodemographic data, maternal medical history, pregnancy and birth history, and infant feeding methods were collected. Infant growth and body composition parameters were obtained at birth and again at 4–6 months of age. For this analysis, infants had to be born term (≥ 37 weeks gestation) and have complete data on body composition available for both the neonatal visit (< 3 days of age) and the infant visit (~ 5 months of age, range 3–7 months). The study was approved by the Colorado Multiple Institutional Review Board and all women provided written informed consent for study participation. The Healthy Start Study was registered at clinicaltrials.gov (NCT02273297).

Mothers self-reported their race/ethnicity, tobacco use, and infant feeding status. Maternal race was categorized as Non-Hispanic White, Hispanic, Non-Hispanic Black, and Other, which included Asian or Pacific Islander and American Indian or Alaskan Native. The total number of breastfeeding months was calculated as previously described [12]. Pre-pregnancy body mass index (BMI), maternal weight gained during pregnancy, gestational age at birth, and infant birthweight were abstracted from medical records.

Infant length was measured supine on a length board to the nearest tenth of a centimeter (cm) by two trained research personnel. Body composition was assessed using air displacement plethysmography (PEA POD, COSMED, Rome, Italy). The PEA POD is a validated instrument that measures total body mass in grams, total body volume, and estimates FM, fat-free mass (FFM), and %FM [15]. Each infant was measured twice, with a third measurement obtained when %FM differed by $> 2.0\%$; the average of the two closest readings was used for analysis. The rate of change for each of the following parameters: length, %FM, FM, and FFM, was calculated as the measurement at ~ 5 months minus the measurement at birth divided by the amount of time between measures in weeks.

Statistical analysis

Descriptive statistics were used to summarize baseline characteristics of the sample as well as body composition parameters at birth

and at 5 months of age stratified by sex. Change in body composition parameters between time points were assessed with a paired t-test. Sex differences were evaluated using t-tests for continuous variables and the Cochran-Mantel-Haenszel test for categorical variables.

Separate general linear univariate multivariable models were fit for each outcome with sex of the offspring as the main predictor. *A priori* covariates included race and the measure of the outcome near birth (i.e. %FM near birth was included as covariate when the rate of change of %FM was the outcome). Although several other covariates are known to affect infant adiposity, including maternal BMI, weight gain during pregnancy, smoking, and infant feeding source, these variables are not associated with infant sex or on the causal pathway between sex and change in body composition so they were not included in the model. We assessed the significance of the covariates using an F test at an α level of 0.05. We examined the studentized residuals to ensure we met the model assumptions of normality and homoscedasticity. All model assumptions were met and there were no overly influential observations. The final model for each outcome included race and the measure of the outcome at birth as covariates. The estimates of association, p-values, and 95% confidence intervals are presented. Statistical analyses were conducted in SAS 9.4 (SAS Institute, Cary, NC, USA).

Results

Out of the full 1410 infant cohort, we excluded infants who did not have outcome measures at both time points ($n = 673$), as well as infants born before 37 weeks' gestation ($n = 11$), infants who had their initial PEA POD measurement after 3 days of age, or second PEA POD more than 209 days after birth ($n = 124$). Therefore, the analytic cohort included the remaining $n = 602$ participants. Maternal and infant characteristics were similar between the full cohort and the analytic cohort, similar to that previously reported [12]. The mean age of the cohort at the first visit was 1.5 days and at the second visit was 5 months. Demographic data were similar between males and females (Table 1). Sex differences in length, FM, FFM, and %FM were present at birth for all parameters ($p < 0.001$, Table 2). As expected, length, FM, FFM, and %FM all increased between birth and 5 months in both males and females ($p < 0.001$ for all, Table 1). At 5 months of age, FFM and %FM remained significantly different between males and females, with relative total body adiposity now 8.7% lower in males compared to females $[(25.47 - 23.25)/25.47, p < 0.0001]$. FM was no longer significantly different between sexes at 5 months of age ($p = 0.12$).

Sex differences in the rate of change per week in body composition parameters are shown in Table 3. On average, after adjusting for race/ethnicity, FFM increased 17 g more per week in males compared to females ($p < 0.0001$, 95% CI: 14, 20), yielding a difference of 410 g of FFM between sexes at 5 months. The rate of FM gain was not significantly

Table 1: Participant demographics (mean $\pm$ SD, n [%]).

	Total sample (n = 602)	Females (n = 306)	Males (n = 296)
Maternal characteristics			
Race			
Hispanic	147 (24%)	74 (24%)	73 (25%)
Non-Hispanic White	341 (57%)	173 (57%)	168 (57%)
Non-Hispanic Black	83 (14%)	42 (14%)	41 (14%)
Other	31 (5%)	17 (6%)	14 (5%)
Pre-pregnancy BMI, kg/m ²	25.84 $\pm$ 6.36	25.68 $\pm$ 6.20	26.00 $\pm$ 6.52
Gestational weight gain, kg	13.65 $\pm$ 6.38	13.65 $\pm$ 6.68	13.65 $\pm$ 6.06
Smoking during pregnancy, n	38 (6%)	18 (6%)	20 (7%)
Infant characteristics			
Gestational age, weeks	39.54 $\pm$ 1.15	39.57 $\pm$ 1.16	39.50 $\pm$ 1.13
Birthweight, g	3277 $\pm$ 426	3220 $\pm$ 410	3330 $\pm$ 428
Age at PEA POD #1, days	1.12 $\pm$ 0.53	1.17 $\pm$ 0.56	1.08 $\pm$ 0.49
Age at PEA POD #2, months	4.94 $\pm$ 0.93	4.98 $\pm$ 0.92	4.89 $\pm$ 0.94
Breastfeeding months	3.48 $\pm$ 1.81	3.47 $\pm$ 1.80	3.49 $\pm$ 1.83

Table 2: Body composition measures (mean $\pm$ SD) at birth and at 5 months by infant sex.

	Total sample (n = 602)	Females (n = 306)	Males (n = 296)	p-Value
Body composition at birth				
Total body mass, kg	3.12 $\pm$ 0.41	3.07 $\pm$ 0.40	3.17 $\pm$ 0.41	0.001
Length, cm	49.18 $\pm$ 2.04	48.75 $\pm$ 1.97	49.63 $\pm$ 2.02	<0.0001
Fat free mass, kg	2.83 $\pm$ 0.33	2.76 $\pm$ 0.31	2.91 $\pm$ 0.33	<0.0001
Fat mass, kg	0.29 $\pm$ 0.14	0.31 $\pm$ 0.14	0.27 $\pm$ 0.13	0.0002
% Fat mass	9.02 $\pm$ 3.72	9.83 $\pm$ 3.85	8.18 $\pm$ 3.39	<0.0001
Body composition at 5 months				
Total body mass, kg	6.74 $\pm$ 0.83	6.57 $\pm$ 0.82	6.92 $\pm$ 0.81	<0.0001
Length, cm	63.74 $\pm$ 2.66	63.09 $\pm$ 2.54	64.44 $\pm$ 2.62	<0.0001
Fat free mass, kg	5.08 $\pm$ 0.57	4.88 $\pm$ 0.53	5.29 $\pm$ 0.54	<0.0001
Fat mass, kg	1.66 $\pm$ 0.50	1.70 $\pm$ 0.50	1.63 $\pm$ 0.50	0.12
% Fat mass	24.38 $\pm$ 5.42	25.47 $\pm$ 5.32	23.25 $\pm$ 5.30	<0.0001

Table 3: Sex differences in the rate of change per week in body composition parameters.

Outcome	β Coefficient (males relative to females)	Standard error	p-Value
Rate of change in percent fat mass, %	-0.09	0.02	0.0004
Rate of change in fat free mass, g/week	16.59	1.65	<0.0001
Rate of change in fat mass, g/week	-1.07	2.00	0.59
Rate of change in length, cm/week	0.05	0.01	<0.0001

different between sexes. For males, %FM increased 0.09 percentage points less per week than females ($p=0.0004$, 95% CI: 0.04, 0.13), which equates to an 8% difference between sexes in adiposity gain during the first 5 months

of life. The length increased 0.05 cm more per week (2.6 cm/year) in males compared to females ($p<0.0001$, 95% CI: 0.04, 0.07) during the first 5 months of life.

Discussion

In this large, ethnically diverse cohort of term infants, we have shown that while sex differences in body composition are already present at birth, the magnitude of these differences increases during the first 5 months of life. The gain in adiposity was 8% lower in males compared to females, and this was due to a greater gain in absolute FFM in males vs. females. We also confirmed that the linear growth velocity is greater in males during this time period. These early differences in growth patterns between males and females are important to consider as we further define the critical role of programming

in the early post-natal period on future risk for adult diseases [16].

Sex differences in adiposity are well recognized in adults, with males having less total adipose tissue but disproportionately more visceral adipose tissue [17]. This sexual dimorphism in body composition has been demonstrated even in pre-pubertal children, although to a much smaller extent than after puberty [1, 18, 19]. Adipose tissue was once thought to be a storage depot for lipids but it is now recognized as an endocrine organ that regulates metabolism [20]. Adipose tissue dysfunction is strongly associated with systemic inflammation, insulin resistance, and cardiovascular disease. Furthermore, accumulation of adiposity in the early infancy period has been associated with later childhood, adolescent, and adult obesity, supporting the model that programming during critical windows of development and plasticity influence lifelong metabolism [21, 22]. Despite this, there has been minimal investigation comparing differences in early growth between sexes, and our understanding of the origins of later sex differences is largely speculative. This study specifically compares longitudinal changes in body composition during this critical time period between males and females.

The mini-puberty period of infancy refers to temporary activation of the hypothalamic-pituitary-gonadal axis and production of sex steroids (testosterone and estrogen) in infancy [7]. The time course is more well-defined in males, peaking around 6–8 weeks of age and typically ending by 4–5 months of age. Testosterone is an anabolic hormone that increases lean mass as well as linear growth velocity. Kiviranta et al. found sex differences in growth velocity in 18,570 infants from the UK, with males growing 2–4 cm/year more than females in the first 6 months of life and no sex differences in growth velocity after 6 months of age [8]. Our data corroborate these findings with a difference between sexes of 2.6 cm/year in the first 5 months of life. These results support the importance of the first few months of life in establishing early sex differences in height. Beyond height, the body composition changes we observed in males in this study are identical to what is expected from testosterone exposure: greater FFM, lower %FM, and higher linear growth velocity compared to females. To confirm these observed changes in body composition parameters were not secondary to the change in length, the analyses were repeated with change in length as a covariate in the model and the outcomes were unchanged. Although we cannot definitively conclude that these differences are due to testosterone, the body composition changes we observed during the time course of the normal mini-puberty period are consistent

with the known effects of testosterone. The implications of the mini-puberty period and its coincidental timing with the hypothesized critical programming in the first months of life are just beginning to be investigated.

The major strengths of this study are the large, ethnically diverse cohort of infants with a rigorous assessment of body composition. Limitations of this study include no measures to directly attribute the observed sex differences to testosterone and data limited at this time to the first 6 months of life (though additional follow-up is underway). In addition, as the PEA POD assesses total body adiposity we are unable to evaluate sex differences in adiposity depots, such as visceral vs. subcutaneous fat, which is recognized to be sex specific and have a strong correlation with cardiometabolic disease states in adults [23]. Despite these limitations, this study is the first to demonstrate that sex differences in body composition increase during these early months of life and provide justification for further investigation into the mechanisms and clinical implications of these findings.

In conclusion, our study confirms there are sex differences emerging in the rate of linear growth during early infancy, and also provides novel data supporting emerging sex differences in body composition. Sex differences in body composition are present at birth but significantly widen over the first 5 months of life. FFM increases significantly more in males than females, while FM gain is similar between sexes resulting in lower adiposity gain in males. Although we cannot confirm this is secondary to the testosterone surge occurring in healthy male infants during this time period with our dataset, this pattern of change in body composition is congruent with what would be expected from testosterone exposure. The finding supports further study of the physical sex differences emerging during the mini-puberty period of infancy that may have lifelong implications in health and disease.

Author contributions: Davis wrote the first draft of the manuscript and each co-author listed contributed to and approved the final manuscript. All authors agree to be accountable for all aspects of the work.

Research funding: This work was funded by NIH/NIDDK Funder Id: <http://dx.doi.org/10.13039/1000000062, R01DK076648> (Dabelea), NIH/OD Funder Id: <http://dx.doi.org/10.13039/1000000052, UH3OD023248> (Dabelea), NIH/NIGMS Funder Id: <http://dx.doi.org/10.13039/1000000057, R01GM121081> (Glueck, Muller, Dabelea), and NIH/NICHD Funder Id: <http://dx.doi.org/10.13039/100009633, K23HD092588> (Davis).

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

Statement of ethics: Parents of participants have given their written informed consent. The study was approved by the research institute's committee on human research. The authors have no ethical conflicts to disclose.

Disclosure statement: The authors have no conflicts of interest to declare.

References

- Hanks LJ, Casazza K, Alvarez JA, Fernandez JR. Does fat fuel the fire: independent and interactive effects of genetic, physiological, and environmental factors on variations in fat deposition and distribution across populations. *J Pediatr Endocrinol Metab* 2010;23:1233–44.
- Butte NF, Hopkinson JM, Wong WW, Smith EO, Ellis KJ. Body composition during the first 2 years of life: an updated reference. *Pediatr Res* 2000;47:578–85.
- Fields DA, Gilchrist JM, Catalano PM, Gianni ML, Roggero PM, et al. Longitudinal body composition data in exclusively breast-fed infants: a multicenter study. *Obesity* 2011;19:1887–91.
- Fomon SJ, Nelson SE. Body composition of the male and female reference infants. *Annu Rev Nutr* 2002;22:1–17.
- Andersen GS, Girma T, Wells JC, Kaestel P, Leventi M, et al. Body composition from birth to 6 mo of age in Ethiopian infants: reference data obtained by air-displacement plethysmography. *Am J Clin Nutr* 2013;98:885–94.
- Jain V, Kurpad AV, Kumar B, Devi S, Sreenivas V, et al. Body composition of term healthy Indian newborns. *Eur J Clin Nutr* 2016;70:488–93.
- Rey RA. Mini-puberty and true puberty: differences in testicular function. *Ann Endocrinol (Paris)* 2014;75:58–63.
- Kiviranta P, Kuiri-Hanninen T, Saari A, Lamidi ML, Dunkel L, et al. Transient postnatal gonadal activation and growth velocity in infancy. *Pediatrics* 2016;138:e2015356.
- Copeland KC, Chernausk S. Mini-puberty and growth. *Pediatrics* 2016;138:e20161301.
- Nohara K, Zhang Y, Waraich RS, Laque A, Tiano JP, et al. Early-life exposure to testosterone programs the hypothalamic melanocortin system. *Endocrinology* 2011;152:1661–9.
- Perng W, Ringham BM, Glueck DH, Sauder KA, Starling AP, et al. An observational cohort study of weight- and length-derived anthropometric indicators with body composition at birth and 5 mo: the Healthy Start study. *Am J Clin Nutr* 2017;106:559–67.
- Sauder KA, Kaar JL, Starling AP, Ringham BM, Glueck DH, et al. Predictors of infant body composition at 5 months of age: the Healthy Start Study. *J Pediatr* 2017;183:94–9.e1.
- Harrod CS, Chasan-Taber L, Reynolds RM, Fingerlin TE, Glueck DH, et al. Physical activity in pregnancy and neonatal body composition: the Healthy Start study. *Obstet Gynecol* 2014;124(2 Pt 1):257–64.
- Harrod CS, Reynolds RM, Chasan-Taber L, Fingerlin TE, Glueck DH, et al. Quantity and timing of maternal prenatal smoking on neonatal body composition: the Healthy Start Study. *J Pediatr* 2014;165:707–12.
- Ma G, Yao M, Liu Y, Lin A, Zou H, et al. Validation of a new pediatric air-displacement plethysmograph for assessing body composition in infants. *Am J Clin Nutr* 2004;79:653–60.
- Koletzko B, Brands B, Chourdakis M, Cramer S, Grote V, et al. The Power of Programming and the EarlyNutrition project: opportunities for health promotion by nutrition during the first thousand days of life and beyond. *Ann Nutr Metab* 2014;64:187–96.
- Palmer BF, Clegg DJ. The sexual dimorphism of obesity. *Mol Cell Endocrinol* 2015;402:113–9.
- Karlsson AK, Kullberg J, Stokland E, Allvin K, Gronowitz E, et al. Measurements of total and regional body composition in pre-school children: a comparison of MRI, DXA, and anthropometric data. *Obesity* 2013;21:1018–24.
- Staiano AE, Broyles ST, Gupta AK, Katzmarzyk PT. Ethnic and sex differences in visceral, subcutaneous, and total body fat in children and adolescents. *Obesity* 2013;21:1251–5.
- Kershaw EE, Flier JS. Adipose tissue as an endocrine organ. *J Clin Endocrinol Metab* 2004;89:2548–56.
- Wells JC, Chomtho S, Fewtrell MS. Programming of body composition by early growth and nutrition. *Proc Nutr Soc* 2007;66:423–34.
- Varvarigou AA. Intrauterine growth restriction as a potential risk factor for disease onset in adulthood. *J Pediatr Endocrinol Metab* 2010;23:215–24.
- Fried SK, Lee MJ, Karastergiou K. Shaping fat distribution: new insights into the molecular determinants of depot- and sex-dependent adipose biology. *Obesity* 2015;23:1345–52.

ORIGINAL ARTICLE

Physical fitness reference standards in European children: the IDEFICS study

P De Miguel-Etayo^{1,2,3,16}, L Gracia-Marco^{1,4,16}, FB Ortega^{5,6}, T Intemann⁷, R Foraita⁷, L Lissner⁸, L Oja⁹, G Barba¹⁰, N Michels¹¹, M Tornaritis¹², D Molnár¹³, Y Pitsiladis¹⁴, W Ahrens^{7,15} and LA Moreno^{1,2} on behalf of the IDEFICS consortium

BACKGROUND/OBJECTIVES: A low fitness status during childhood and adolescence is associated with important health-related outcomes, such as increased future risk for obesity and cardiovascular diseases, impaired skeletal health, reduced quality of life and poor mental health. Fitness reference values for adolescents from different countries have been published, but there is a scarcity of reference values for pre-pubertal children in Europe, using harmonised measures of fitness in the literature. The IDEFICS study offers a good opportunity to establish normative values of a large set of fitness components from eight European countries using common and well-standardised methods in a large sample of children. Therefore, the aim of this study is to report sex- and age-specific fitness reference standards in European children.

SUBJECTS/METHODS: Children (10 302) aged 6–10.9 years (50.7% girls) were examined. The test battery included: the flamingo balance test, back-saver sit-and-reach test (flexibility), handgrip strength test, standing long jump test (lower-limb explosive strength) and 40-m sprint test (speed). Moreover, cardiorespiratory fitness was assessed by a 20-m shuttle run test. Percentile curves for the 1st, 3rd, 10th, 25th, 50th, 75th, 90th, 97th and 99th percentiles were calculated using the General Additive Model for Location Scale and Shape (GAMLSS).

RESULTS: Our results show that boys performed better than girls in speed, lower- and upper-limb strength and cardiorespiratory fitness, and girls performed better in balance and flexibility. Older children performed better than younger children, except for cardiorespiratory fitness in boys and flexibility in girls.

CONCLUSIONS: Our results provide for the first time sex- and age-specific physical fitness reference standards in European children aged 6–10.9 years.

International Journal of Obesity (2014) 38, S57–S66; doi:10.1038/ijo.2014.136

INTRODUCTION

Physical fitness (hereinafter called just fitness) has been related to a person's ability to perform physical activities that require aerobic capacity, endurance, strength or flexibility, which seems to be linked to inherited and environmental factors.¹ Fitness has been considered a powerful marker of health, both, in childhood and in adulthood,² independent of physical activity.^{3,4} A low fitness status during childhood and adolescence is associated with important health-related outcomes, such as increased future risk for obesity⁵ and cardiovascular diseases,⁶ impaired skeletal health,⁷ reduced quality of life⁸ and poor mental health.⁹ In spite of the healthy benefits of a high fitness, children and adolescents performance in fitness tests has declined over the last three decades.¹⁰

The most frequently evaluated fitness component is cardiorespiratory fitness (CRF).¹¹ However, in the last years other fitness components such as flexibility, muscular fitness and speed/agility were evaluated and associated with health outcomes.⁹

Nowadays, fitness reference values in adolescents from different countries in America,^{12–14} Asia,¹⁰ Africa¹¹ and Europe⁹ have been

published. However, there is a scarcity of reference values for pre-pubertal children in Europe and in other continents, using harmonised measures of fitness in the literature.¹⁴ Reference values are necessary to classify children and to monitor the fitness status of the population. The IDEFICS study (Identification and prevention of dietary and lifestyle-induced health effects in children and infants)¹⁵ offers a good opportunity to establish normative values for a large set of fitness components using common and well-standardised methods in a large sample of children from eight European countries. Therefore, the aim of this study is to provide sex- and age-specific fitness reference standards for European children aged 6–9 years.

MATERIALS AND METHODS

Study design

The IDEFICS study¹⁵ is a prospective cohort study with an embedded controlled intervention aiming to prevent childhood obesity in a community-oriented approach. IDEFICS is a multi-centre study on lifestyle

¹GENUD (Growth, Exercise, NUTrition and Development) Research Group, Universidad de Zaragoza, Zaragoza, Spain; ²Department of Psychiatry and Nursing, Faculty of Health Sciences, University of Zaragoza, Zaragoza, Spain; ³Department of Paediatrics, Faculty of Medicine, Universidad de Zaragoza, Zaragoza, Spain; ⁴CHERC (Children's Health and Exercise Research Centre), College of Life and Environmental Sciences, Sport and Health Sciences, University of Exeter, Exeter, UK; ⁵PROFITH (PROmoting FITness and Health through physical activity) Research Group, School of Sports Science, University of Granada, Granada, Spain; ⁶Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden; ⁷Leibniz Institute for Prevention Research and Epidemiology—BIPS, Bremen, Germany; ⁸Department of Public Health and Community Medicine, University of Gothenburg, Gothenburg, Sweden; ⁹National Institute for Health Development, Tallinn, Estonia; ¹⁰Epidemiology and Population Genetics, Institute of Food Science, National Research Council, Avellino, Italy; ¹¹Department of Public Health, Ghent University, Ghent, Belgium; ¹²Research and Education Institute for Child Health, Strovolos, Cyprus; ¹³Department of Pediatrics, University of Pécs Medical School, Pécs, Hungary; ¹⁴Centre for Sport and Exercise Science and Medicine (SESAME), University of Brighton, Eastbourne, UK and ¹⁵Institute of Statistics, Faculty of Mathematics and Computer Science, University of Bremen, Bremen, Germany. Correspondence: P De Miguel-Etayo, GENUD (Growth, Exercise, NUTrition and Development) Research Group, Universidad de Zaragoza, C/Pedro Cerbuna, 12, Zaragoza 50009, Spain.

E-mail: pilardm@unizar.es

¹⁶These authors contributed equally to this work.

and nutrition among children aged 2–9 years from eight European countries (Sweden, Germany, Hungary, Italy, Cyprus, Spain, Belgium and Estonia).

In the first survey, data collection took place from September 2007 to June 2008 (T0, baseline survey); 2 years later the follow-up survey was conducted from September 2009 to May 2010 (T1, follow-up survey), where the present analysis is based on cross-sectional data only. A detailed description of IDEFICS sampling and recruitment approaches, standardisation and harmonisation process, data collection, analysis strategies, quality control measures and inclusion/exclusion criteria have been already published.¹⁵ The study was approved by the Research Ethics Committees of each institution and region involved. Parents signed an informed consent.

Study population

A cohort of 18 745 children aged 2–10.9 years was established including all children recruited at baseline and children newly recruited at the first

follow-up. Fitness was measured in children older than 6 years ($N = 10\,302$, 50.7% girls) fulfilling the inclusion criteria (having complete data on weight, height, age and sex). Because not all children took part in all components of the physical fitness battery, analysis groups and sample sizes vary for the different physical fitness tests (Figure 1). We compared these varying analysis groups with the overall study population of children older than 6 years. Since the groups of children who participated in the back-saver sit-and-reach test, the standing long jump or the handgrip strength test were nearly identical, we considered these children as one analysis group. The second analysis group consisted of children who participated in the 40-m sprint test; the third group of children who participated in the Flamingo balance test; and the children who conducted the shuttle run test formed the fourth group. Mean age varied between 7.6 and 7.7 years, and the percentage of boys was highest in the total study population with 49.4% and smallest (47.3%) in the third analysis group (Flamingo balance test). The prevalence of overweight and obese children



Figure 1. Flow chart of the population involved in this study from the IDEFICS study.

was highest (25.3%) for children of group 1 compared with 24.5% in the total study population and smallest (20.9%) in the children participating in the shuttle run test. The most pronounced differences were seen with respect to the ISCED (International Standard Classification of Education) level where the maximum level of both parents was considered. The distribution of ISCED levels was rather similar in the overall study population, analysis groups 1 and 3, but fewer children from families with ISCED level 0–2 participated in the 40-m sprint (4.5%; analysis group 2) and in the shuttle run test (6.8%; analysis group 4) as compared with 10.5% in the overall study population, 10.3% in group 1 and 10.0% in group 3.

Physical fitness

Components of the fitness tests were mostly adapted from the ALPHA health-related fitness test battery, and based on a published review their reliability has been shown in children and adolescents.^{9,16–18} The test battery thus included: the flamingo balance test, back-saver sit-and-reach test, handgrip strength test, standing long jump test and 40-m sprint test. Moreover, CRF was assessed by the 20-m shuttle run test (20mSRT).¹⁹

The protocols used for fitness testing are described in detail below:

The flamingo balance test measures the ability to balance successfully on a single leg. The child has to bend his/her free leg backwards and grip the back foot with his/her hand on the same side, and stand like this for 1 min. The child is given one try before to become familiar with the test. Then the number of attempts needed to stand on one leg for 1 min is counted for each leg. Children were excluded if they had put down the other foot 15 times or more within the first 30 s. The test score is calculated as the sum of attempts with both legs; lower scores indicate better performance.

The back-saver sit-and-reach test measures the flexibility of the hamstring muscles and it was conducted in the FITNESSGRAM battery.²⁰ The test is performed with a standard box with a scale on the top. The participant is required to sit with the untested leg bent at the knee; the tested leg is placed straight with the foot placed against the box. In the back-saver sit-and-reach test, only one leg was evaluated at a time. The participant slowly reached forward as far as possible. The back-saver sit-and-reach test is similar to the traditional sit-and-reach test, except that the measurement is performed on one side at a time, so a specific score is obtained from each side. The score is calculated as the average of both sides; higher scores indicate better performance.

The handgrip strength test measures the maximal isometric force that can be generated mainly by the forearm. The child stays in a standard bipedal position with the arms in complete extension holding the dynamometer (TKK 5101; Takei, Tokyo, Japan) without touching any part of the body with it. The dynamometer is adjusted to sex and hand size for each child.²¹ The measurement scale starts with 5 kg. Children who did not

reach this minimum ($N=5$; 0.1%) were scored as 2.5 kg (average 0–5 kg). The score is calculated as the average of right and left handgrip strength; higher scores indicate better performance.

The standing long jump test²² assesses lower-limb explosive strength. The child jumps as far as possible off the stand, trying to land with both feet together and maintaining the equilibrium once landed (it was not allowed to put the hands on the floor). The score was obtained by measuring the distance between the last heel-mark and the take-off line. Two tries were allowed and the best score was retained. Higher scores indicate better performance.

The 40-m sprint test measures the maximum running speed of the child. This test is carried out along 40 m delimited by five marker cones aligned, within a distance of 10 m between each two neighbouring cones. With 3-m distance, five more marker cones are placed in parallel marking the running track. The child is instructed to run as quick as he/she can, after the starting signal. Two tries were allowed, the best score was retained. In this test lower scores indicate better performance.

The 20mSRT estimates the CRF (aerobic capacity). Children run back and forth between two lines 20-m apart following beep signals played from a pre-recorded CD. The test is finished when the child stops owing to fatigue or when he/she does not reach the line in time with the audio signal on two consecutive occasions. During the fieldwork, 20mSRT was performed using four different versions: multistage fitness test, sports coach UK (applied in Germany, Estonia and Cyprus); Leger test, CAEP Quebec Faca (applied in Spain and Hungary); multistage fitness test (applied in Sweden) and Uithouding shuttle run test (applied in Belgium). During the fieldwork, 20mSRT was assessed in shuttles. Results were unified according to the Leger test protocol. The initial speed in the Leger test starts in 8.5 km h^{-1} , with progressive increases of 0.5 km h^{-1} . Taking into account the speed in the other three protocols, we estimated the equivalent shuttles for every protocol. Shuttles were converted to stages in order to calculate maximal oxygen consumption ($\text{VO}_{2\text{max}}$) using Leger's equation.¹⁹ A greater number of shuttles indicate better performance. This test was not performed in Italian children ($N=2440$; 51.8% girls).

Anthropometric measurements

International guidelines for anthropometry in children were used in the IDEFICS study.^{23,24} Body weight (kg) and height (cm) were measured in barefoot children, clothed in underwear, using an adapted version of electronic scale TANITA BC 420 SMA (Tanita Europe BV, Amsterdam, The Netherlands), precision 100 g, range 0–150 kg and a portable stadiometer (seca 225, seca, Birmingham, UK), precision 0.1 cm, range 70–200 cm, respectively. Body mass index was calculated as body weight (kg) divided by height (m) squared.

Table 1. Chosen GAMLSS models to calculate the physical fitness reference values

Variable	Sex	Model distribution	Parameters			
			μ	$\log(\sigma)$	ν	$\log(\tau)$
Back-saver sit-and-reach test (cm)	Girls	BCCG	1	Age	1	—
	Boys	BCCG	Age	Age	1	—
Handgrip strength test (kg)	Girls	BCCG	Age	1	Age	—
	Boys	BCCG	Age	1	1	—
Standing long jump test (cm)	Girls	BCCG	Age	Age	1	—
	Boys	BCCG	Age	Age	1	—
40-m sprint test (s)	Girls	BCPE	Age	1	1	1
	Boys	BCPE	Age	1	1	1
20-m shuttle run test (shuttles)	Girls	BCPE	Age	1	1	1
	Boys	BCPE	$\log(\mu)$	$\log(\sigma)$	$\text{logit}(\nu)$	—
Flamingo balance test (attempts)	Girls	DEL	Age	Age	Age	—
	Boys	DEL	$\log(\mu)$	$\log(\sigma)$	—	—
20-m shuttle run test ($\text{VO}_{2\text{max}}$, $\text{ml kg}^{-1} \text{ min}^{-1}$)	Girls	IG	Age	Age	—	—
	Boys	IG	Age	Age	—	—

Abbreviations: BCCG, Box–Cox Cole and Green; BCPE, Box–Cox power exponential; DEL, Delaporte distribution; GAMLSS, General Additive Model for Location Scale and Shape; IG, inverse Gaussian.

Table 2. Percentiles of coordination/equilibrium; flexibility and speed/agility calculated with GAMLSS in normal-weight children

	Percentile for girls										Percentile for boys									
	Age	1	3	10	25	50	75	90	97	99	Age	1	3	10	25	50	75	90	97	99
Flamingo balance test (attempts)																				
6- < 6.5	1	1	1	3	4	8	13	20	28	36	6- < 6.5	1	2	4	6	11	17	24	33	41
6.5- < 7	0	1	2	3	6	10	17	26	34		6.5- < 7	1	1	3	5	9	14	21	30	38
7- < 7.5	0	1	2	3	5	8	14	23	32		7- < 7.5	0	1	2	4	7	12	19	27	36
7.5- < 8	0	0	1	2	4	7	12	20	29		7.5- < 8	0	1	2	3	5	10	16	25	33
8- < 8.5	0	0	1	2	3	5	9	17	25		8- < 8.5	0	1	1	3	5	8	13	22	30
8.5- < 9	0	0	1	2	3	5	7	14	22		8.5- < 9	0	0	1	2	4	6	11	19	27
Back-saver sit-an-reach test (cm)											Back-saver sit-and-reach test (cm)									
6- < 6.5	10.5	12.9	16.0	18.9	21.9	24.8	27.3	29.7	31.4		6- < 6.5	7.5	10.1	13.3	16.5	19.8	23.1	25.9	28.6	30.6
6.5- < 7	9.7	12.3	15.6	18.7	21.9	25.0	27.6	30.1	31.9		6.5- < 7	7.1	9.7	13.0	16.2	19.6	22.9	25.8	28.6	30.6
7- < 7.5	8.9	11.7	15.2	18.5	21.9	25.1	27.9	30.6	32.5		7- < 7.5	6.6	9.2	12.7	15.9	19.4	22.8	25.7	28.5	30.6
7.5- < 8	8.1	11.1	14.8	18.3	21.9	25.3	28.3	31.1	33.1		7.5- < 8	6.2	8.8	12.3	15.7	19.2	22.6	25.6	28.5	30.6
8- < 8.5	7.3	10.4	14.4	18.1	21.9	25.5	28.6	31.6	33.7		8- < 8.5	5.7	8.4	12.0	15.4	19.0	22.5	25.5	28.4	30.6
8.5- < 9	6.5	9.7	13.9	17.8	21.9	25.7	29.0	32.1	34.4		8.5- < 9	5.3	8.1	11.6	15.1	18.8	22.4	25.4	28.4	30.6
40-m sprint (s)											40-m sprint (s)									
6- < 6.5	8.1	8.5	8.7	9.2	9.9	10.7	11.5	12.4	13.1		6- < 6.5	7.6	7.8	8.3	8.8	9.4	10.2	11.0	11.8	12.6
6.5- < 7	7.9	8.3	8.5	9.0	9.6	10.4	11.2	12.1	12.8		6.5- < 7	7.4	7.7	8.1	8.6	9.2	10.0	10.8	11.6	12.3
7- < 7.5	7.7	8.0	8.3	8.7	9.4	10.1	10.9	11.8	12.5		7- < 7.5	7.2	7.5	7.9	8.4	9.0	9.8	10.5	11.3	12.0
7.5- < 8	7.5	7.8	8.1	8.5	9.1	9.9	10.6	11.4	12.1		7.5- < 8	7.1	7.3	7.7	8.2	8.8	9.5	10.3	11.1	11.7
8- < 8.5	7.3	7.6	7.8	8.3	8.9	9.6	10.3	11.1	11.8		8- < 8.5	6.9	7.2	7.5	8.0	8.6	9.3	10.0	10.8	11.4
8.5- < 9	7.1	7.4	7.6	8.0	8.6	9.3	10.0	10.8	11.5		8.5- < 9	6.7	7.0	7.4	7.8	8.4	9.1	9.8	10.5	11.2

Abbreviation: GAMLSS, General Additive Model for Location Scale and Shape.²⁵ The values corresponded to the 1st, 3rd, 10th, 25th, 50th, 75th, 90th, 97th and 99th age- and sex-specific percentiles.

Table 3. Percentiles of upper-limb maximal strength and lower-limb explosive strength calculated with GAMLSS in normal-weight children

	Percentile for girls										Percentile for boys									
	Age	1	3	10	25	50	75	90	97	99	Age	1	3	10	25	50	75	90	97	99
Handgrip strength (kg)																				
	6- < 6.5	5.0	5.5	6.2	7.0	8.1	9.3	10.6	12.1	13.2	6- < 6.5	5.4	6.0	6.9	7.9	9.1	10.4	11.7	13.0	14.1
	6.5- < 7	5.4	6.0	6.8	7.8	9.0	10.3	11.7	13.1	14.4	6.5- < 7	5.9	6.6	7.6	8.7	10.0	11.4	12.8	14.3	15.4
	7- < 7.5	5.8	6.4	7.4	8.5	9.8	11.2	12.7	14.2	15.4	7- < 7.5	6.4	7.2	8.3	9.4	10.9	12.4	13.9	15.5	16.8
	7.5- < 8	6.1	6.9	8.0	9.2	10.6	12.2	13.7	15.3	16.5	7.5- < 8	7.0	7.8	8.9	10.2	11.7	13.4	15.1	16.8	18.1
	8- < 8.5	6.5	7.3	8.6	9.9	11.5	13.1	14.7	16.3	17.6	8- < 8.5	7.5	8.3	9.6	11.0	12.6	14.4	16.2	18.0	19.5
	8.5- < 9	6.7	7.7	9.1	10.6	12.3	14.1	15.7	17.4	18.6	8.5- < 9	8.0	8.9	10.3	11.7	13.5	15.4	17.3	19.3	20.8
Standing long jump test (cm)																				
	6- < 6.5	46.7	56.6	69.1	81.1	93.8	106.1	116.7	127.0	134.4	6- < 6.5	52.0	63.5	77.3	90.0	103.0	115.0	125.3	134.9	141.8
	6.5- < 7	51.1	61.0	73.6	85.7	98.6	111.1	121.9	132.3	139.9	6.5- < 7	56.6	68.1	82.1	94.9	108.0	120.3	130.7	140.6	147.6
	7- < 7.5	55.6	65.5	78.2	90.4	103.5	116.0	127.0	137.6	145.2	7- < 7.5	61.3	72.8	86.8	99.8	113.1	125.6	136.2	146.2	153.4
	7.5- < 8	60.2	70.1	82.8	95.2	108.3	121.0	132.1	142.7	150.5	7.5- < 8	66.1	77.6	91.7	104.7	118.2	130.8	141.6	151.7	159.0
	8- < 8.5	64.9	74.8	87.5	99.9	113.1	125.9	137.1	147.9	155.7	8- < 8.5	71.0	82.4	96.5	109.7	123.3	136.0	146.9	157.2	164.6
	8.5- < 9	69.6	79.5	92.3	104.7	118.0	130.8	142.1	152.9	160.8	8.5- < 9	75.9	87.3	101.5	114.6	128.3	141.2	152.2	162.6	170.1

Abbreviation: GAMLSS, General Additive Model for Location Scale and Shape.²⁵ The values corresponded to the 1st, 3rd, 10th, 25th, 50th, 75th, 90th, 97th and 99th age- and sex-specific percentiles.**Table 4.** Percentiles of cardiorespiratory fitness: 20-m shuttle run test (shuttles and $\text{VO}_{2\text{max}}$ ($\text{ml kg}^{-1} \text{min}^{-1}$)) calculated with GAMLSS in normal-weight children

	Percentile for girls										Percentile for boys									
	Age	1	3	10	25	50	75	90	97	99	Age	1	3	10	25	50	75	90	97	99
20mSRT (shuttles)																				
	6- < 6.5	3.9	4.9	6.5	8.6	11.8	15.9	20.7	26.7	32.1	6- < 6.5	3.5	4.6	6.6	9.4	13.7	19.3	25.3	32.3	38.2
	6.5- < 7	4.4	5.4	7.3	9.7	13.2	17.8	23.2	29.9	35.9	6.5- < 7	4.0	5.3	7.5	10.7	15.6	21.9	28.8	36.8	43.5
	7- < 7.5	4.8	6.0	8.1	10.7	14.6	19.8	25.7	33.1	39.8	7- < 7.5	4.5	5.9	8.5	12.0	17.4	24.6	32.3	41.2	48.7
	7.5- < 8	5.3	6.6	8.8	11.8	16.0	21.7	28.2	36.3	43.7	7.5- < 8	4.9	6.5	9.4	13.3	19.3	27.2	35.8	45.7	54.0
	8- < 8.5	5.8	7.2	9.6	12.8	17.5	23.6	30.7	39.6	47.6	8- < 8.5	5.4	7.2	10.3	14.6	21.2	29.9	39.3	50.1	59.2
	8.5- < 9	6.2	7.8	10.4	13.9	18.9	25.5	33.2	42.8	51.5	8.5- < 9	5.9	7.8	11.2	15.9	23.1	32.5	42.8	54.5	64.5
20mSRT ($\text{VO}_{2\text{max}}$ ($\text{ml kg}^{-1} \text{min}^{-1}$))																				
	6- < 6.5	42.9	43.7	44.9	46.0	47.4	48.7	50.0	51.3	52.3	6- < 6.5	42.8	43.8	45.1	46.5	48.1	50.6	51.2	52.7	53.9
	6.5- < 7	42.1	43.0	44.2	45.5	47.0	48.5	49.9	51.3	52.4	6.5- < 7	42.1	43.1	44.6	46.1	47.8	50.6	51.3	53.0	54.3
	7- < 7.5	41.2	42.2	43.5	45.0	46.6	48.3	49.8	51.4	52.7	7- < 7.5	41.3	42.4	44.0	45.6	47.5	50.6	51.4	53.3	54.7
	7.5- < 8	40.3	41.4	42.8	44.4	46.2	48.1	49.8	51.6	53.0	7.5- < 8	40.4	41.7	43.4	45.2	47.3	50.7	51.5	53.6	55.2
	8- < 8.5	39.3	40.5	42.1	43.8	45.8	47.9	49.8	51.8	53.4	8- < 8.5	39.5	40.9	42.7	44.7	47.0	50.8	51.7	54.0	55.9
	8.5- < 9	38.3	39.6	41.3	43.2	45.4	47.7	49.9	52.1	53.9	8.5- < 9	38.6	40.0	42.0	44.2	46.7	50.9	51.9	54.5	56.6

Abbreviation: GAMLSS, General Additive Model for Location Scale and Shape.²⁵ 20mSRT, 20-m shuttle run test. The values corresponded to the 1st, 3rd, 10th, 25th, 50th, 75th, 90th, 97th and 99th age- and sex-specific percentiles.

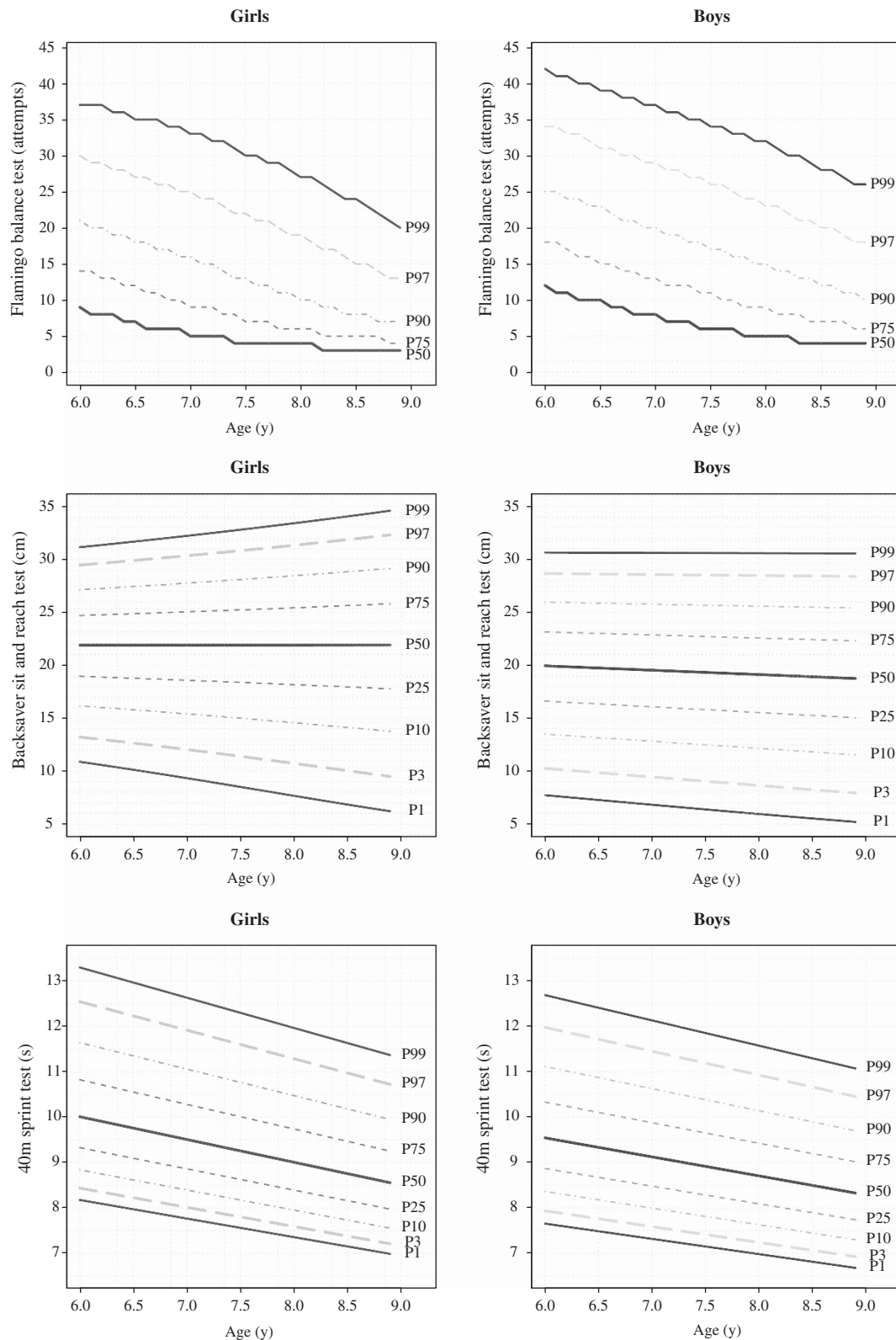


Figure 2. Percentile curves of the flamingo balance test (attempts), back-saver sit-and-reach test (cm) and 40-m sprint test (s) at median age in normal-weight children.

Statistical analysis

We calculated percentile curves of the physical fitness outcome variables as a function of the covariate age stratified by sex using the General Additive Model for Location Scale and Shape (GAMLSS) method. The

GAMLSS method is an extension of the LMS method that models three parameters depending on one explanatory variable: μ accounts for the median of the outcome variable and the coefficient of variation, σ accounts for the variation around the mean and adjusts for non-uniform

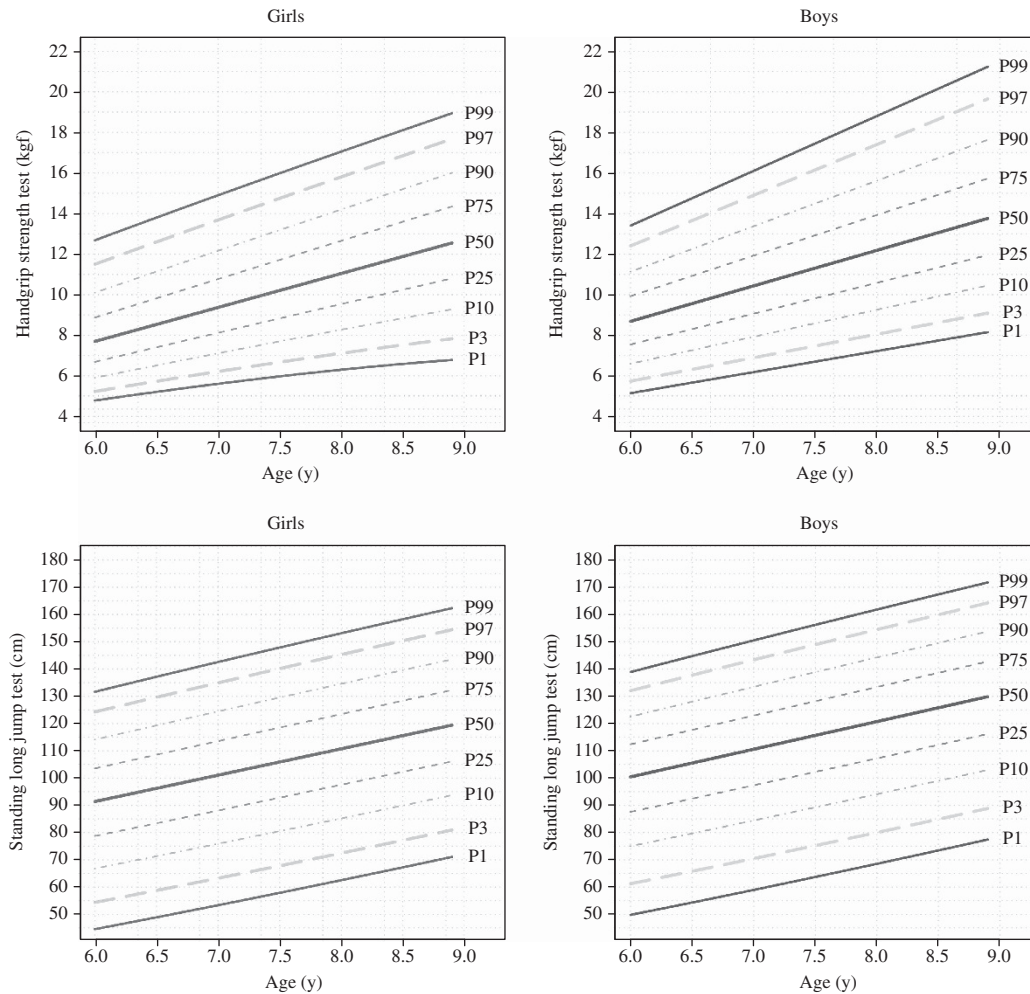


Figure 3. Percentile curves of handgrip strength test (kg) and standing long jump test (cm) at median age in normal-weight children.

dispersion, whereas the skewness (L) accounts for the deviation from a normal distribution using a Box–Cox transformation. The GAMLSS method is able to particularly model the kurtosis using other distributions and to include > 1 covariate. We used the *gamlss* package (version 4.2-6)²⁵ of the statistical software *R* (version 3.0.1).²⁶ Different distributions, that is, the Box–Cox power exponential (BCPE), Box–Cox Cole and Green, inverse Gaussian and Delaporte distribution were fitted to the observed distribution of the physical fitness outcome variables. Moreover, the influence of age on the parameters of the considered distributions were modelled either as a constant, as a linear function or as a cubic spline. Goodness of fit was assessed by the Bayesian Information Criterion and Q–Q plots to select the final model including the fitted distribution of the physical fitness outcome variables. Worm plots were used as a diagnostic tool to assess whether adjustment for kurtosis was required.²⁷ Finally, percentile curves for the 1st, 3rd, 10th, 25th, 50th, 75th, 90th, 97th and 99th percentiles were calculated based on the model that showed the best goodness of fit.^{25,28} For the Flamingo balance test only the 50th, 75th, 90th, 97th and 99th percentile curves are presented, because lower percentile curves of the best model did not correspond to the percentage of values below the percentile curves, that is, for example, instead of the targeted 25% of the girls' values there were only 12.6% of the girls' values below the 25th percentile curve. Similar results were observed for the 3rd, 10th and 25th percentile curves of the flamingo balance test in boys and girls.

The chosen GAMLSS models in boys and girls are listed in Table 1; for example, the best model for the 40-m sprint test in boys and girls was achieved with a Box–Cox power exponential distribution where the four parameters were modelled as follows: the location parameter μ linearly, the scale parameter $\log(\sigma)$ and the shape parameters ν and $\log(\tau)$ as

constants. Taking account the discrete distribution of the data from the flamingo balance test in contrast to all other models, the discrete Delaporte distribution was chosen.

RESULTS

Tables 2–4 and Figures 2–4 show the sex- and age-specific reference values (P_1 , P_3 , P_{10} , P_{25} , P_{50} , P_{75} , P_{90} , P_{97} and P_{99}) for the different fitness tests in European children.

Boys performed better than girls in speed/agility (40-m sprint), muscular strength (handgrip strength and standing long jump) and CRF (20mSRT; shuttles and VO_{2max}). Nevertheless, girls performed better than boys in balance (flamingo) and flexibility (back-saver sit-and-reach).

Older children performed better than younger children in all tests, except $\leq P_{50}$ in back-saver sit-and-reach (in both sexes) and $< P_{90}$ in 20mSRT (VO_{2max} in both sexes).

Finally, for all fitness tests, the range of fitness levels between P_3 and P_{99} is wider for a given age than the range of fitness levels across age groups and between sexes.

DISCUSSION

The main findings of this study are (1) boys performed better than girls in speed, lower- and upper-limb strength and CRF; (2) girls performed better in balance and flexibility; (3) overall, older children performed better than younger children; and (4) the

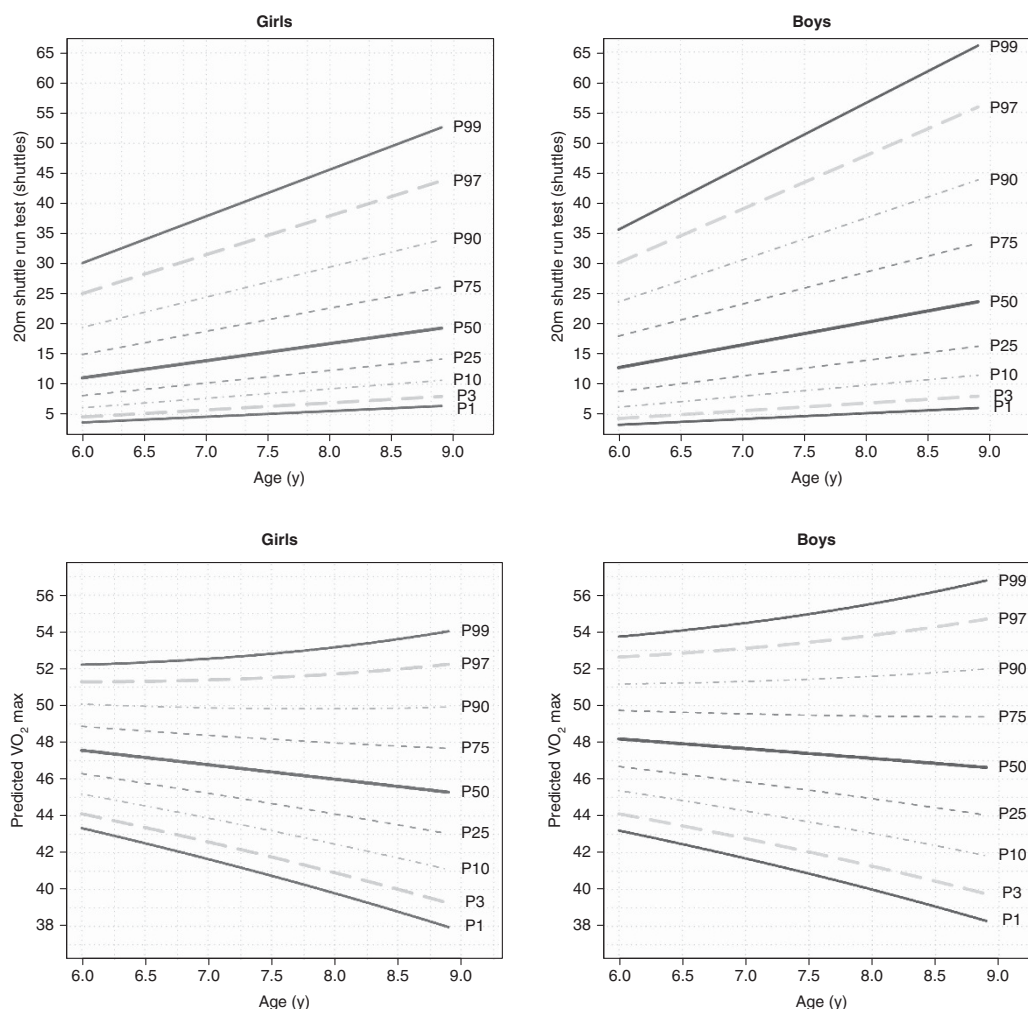


Figure 4. Percentile curves of the 20-m shuttle run test (20mSRT) estimates the CRF (aerobic capacity) at median age in normal-weight children.

range of fitness levels between P_3 and P_{99} is wider for a given age than the range of fitness levels across age groups and between sexes.

The majority of the fitness tests used in the IDEFICS study was selected based on the scientific evidence that showed associations with different health outcomes.²⁹ These tests have also been shown to be valid, reliable and feasible for health-monitoring purposes at the population level.¹⁶ Fitness has been identified as a predictor of morbidity and mortality for all causes.^{30–33} In this regard, various studies have shown that low fitness in children and adolescents is associated with adiposity,^{7,34,35} cardiovascular risk factors^{36,37} and skeletal³⁸ and mental health.³⁹

In 2009, Ruiz *et al.*⁹ systematically reviewed whether fitness in childhood was a predictor of cardiovascular disease risk factors, events and syndromes, quality of life and low back pain later in life. The authors found moderate evidence for the association between changes in CRF and cardiovascular risk factors, and between CRF and the risk of developing metabolic syndrome and arterial stiffness. In addition, CRF has also been widely studied in relation to metabolic risk and adiposity.^{36,37,38} Interestingly, handgrip strength in Swedish male adolescents (aged 16–19 years) has been identified as a risk factor for the major causes of death in adulthood (< 55 years).⁶ Moreover, results from the Amsterdam Growth and Health Study, following adolescents aged 13 years until the age of 27 years, showed that the longitudinal

improvements in VO_{2max} were related to a healthy cardiovascular risk profile.⁴¹

To the best of our knowledge, our study is the first in providing cutoff values of sex- and age-specific fitness for pre-pubertal European children. These values may be useful in identifying children being at higher risk for developing unfavourable health outcomes owing to their low fitness level. Previously, Casajus *et al.*⁴² published physical fitness levels in children from Aragon (Spain) aged 7–12 years. Our results are roughly comparable to these with respect to flexibility (they used sit-and-reach), muscular strength (handgrip strength and standing long jump) and CRF (20mSRT, VO_{2max}) only for those children aged 7–9 years. Similar values were observed in both studies for these fitness tests in boys and girls. For example, P_{50} of sit-and-reach was 17.3 cm vs 19.4 cm for Spanish and European boys, respectively, and 19.3 cm vs 21.9 cm for Spanish and European girls, respectively. P_{50} of handgrip as average of the right and left side was 12.7 kg vs 11.8 kg for Spanish and European boys, respectively, and 11.6 kg vs 10.8 kg for Spanish and European girls, respectively. P_5 of standing long jump was 117.2 cm vs 120.5 cm for Spanish and European boys, respectively, and 104.5 cm vs 110.7 cm for Spanish and European girls, respectively. Finally, P_{50} of 20mSRT was 48.1 ml $kg^{-1} min^{-1}$ vs 47.3 ml $kg^{-1} min^{-1}$ for Spanish and European boys, respectively, and 46.7 ml $kg^{-1} min^{-1}$ vs 46.2 ml $kg^{-1} min^{-1}$ for Spanish and European girls, respectively.

Castro-Pinero *et al.*⁴⁴ published percentile values for muscular strength (standing long jump)⁴³ and CRF (20mSRT, stages) in children and adolescents from Cadiz (Spain) aged 6–17 years. Our results are also comparable to those from Castro-Pinero *et al.*^{43,44} for those children aged 6–9 years. For example, P_{50} of standing long jump was 111.6 cm vs 120.5 cm for Spanish and European boys, respectively, and 114.8 cm vs 110.7 cm for Spanish and European girls, respectively. In addition, P_{50} of 20mSRT (stage) was 2.7 vs 2.0 for Spanish and European boys, respectively, and 1.9 vs 2.0 for Spanish and European girls, respectively.

In adolescent populations, Ortega *et al.*⁴⁵ first published European fitness reference values for 12.5–17.5-year-old youths from 10 cities (HELENA study); reporting sex- and age-specific physical fitness levels. In this line, Haugen *et al.*⁴⁶ recently reported normative fitness data for Norwegian 13–15-year-old adolescents. A systematic review¹⁰ identified Australian studies reporting normative fitness data for children and adolescents aged 9–17 years. In addition, Tremblay *et al.*¹⁴ reported normative data for aerobic fitness, flexibility and muscular strength for Canadian 6–19-year-old youths; however, only the 50th percentile was reported. The present results are comparable to those obtained by Tremblay *et al.*¹⁴ for flexibility (sit-and-reach) and muscular fitness (handgrip strength) in Canadian children (for the age range 6–10 years). For these fitness dimensions, European children performed worse than the Canadian ones. For example, in the study of Tremblay *et al.*¹⁴ the performance in sit-and-reach at P_{50} of 6–10-year-old children was 25 cm and 30 cm in boys and girls, respectively. In the present study, the performance in back-saver sit-and-reach at P_{50} of 6–9-year-old children was 19.4 cm and 21.9 cm for boys and girls, respectively. In addition, in the study of Tremblay *et al.*¹⁴ the performance in handgrip strength at P_{50} was 25 kg and 22 kg in boys and girls, respectively, calculated as the sum of the best right- and left-hand attempt. In our study, the performance in handgrip test at P_{50} of 6–9-year-old children was 23.2 kg and 21 kg for boys and girls, respectively. The differences observed in flexibility might be explained by slightly different methodologies. Tremblay *et al.*¹⁴ measured flexibility by using sit-and-reach and using the best performance of two attempts. In our study, the back-saver sit-and-reach values were calculated as the average of the left and right sides.

There is scarcity of data on reference standards of fitness for children. Hence, the presented European data are relevant in various respects. A fitness level below P_5 may be considered as potentially pathologic since, for example, low scores on CRF and handgrip tests are associated with cardiovascular risk. Therefore, it may be recommended to monitor children with a fitness level below this percentile for cardiovascular risk markers.^{47,48} For a practical use of these data, schools, sport clubs and so on may consider to take into account a Likert type scale to classify children's performance X , as follows: very poor ($X < P_{10}$); poor ($P_{10} \leq X < P_{25}$); medium ($P_{25} \leq X < P_{75}$); good ($P_{75} \leq X < P_{90}$); and very good ($X \geq P_{95}$).

Some limitations should be considered when interpreting the findings in this study. As already mentioned before, the study was not planned to be representative with respect to the broad range of variables that we investigated, but given the population-based approach our study sample should be considered as an unselected population. As the children were free to opt out of examination modules, the analysis groups and the sample sizes of children participating in the various components of the test battery to assess physical fitness varied from test to test. However, comparing the respective analysis groups with the overall sample of 6–10.9-year-old children, no major differences were detected with respect to mean age, sex distribution and prevalence of overweight and obesity. It would have been interesting to investigate whether the reliability of the fitness tests differs in apparently pre-pubertal children and adolescents. However, this was not feasible within our IDEFICS sample but may be investigated in later follow-ups of the cohort. The main strengths

of this study are the large and heterogeneous sample of European children, the standardised use of well-known and validated health-related fitness tests and a strong statistical method to obtain normative values of fitness tests.

In summary, our results provide sex- and age-specific physical fitness reference standards in European children. These data in children aged 6.0–9.9 years complement the study published by Ortega *et al.*⁴⁵ in adolescents aged 12.5–17.5 years. Unfortunately, there is a gap between 9.9 and 12.4 years without updated fitness reference data at the European level, which has to be filled. In the meanwhile, country-specific data such as those commented above should be used in children of this age.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work has been done as part of the IDEFICS Study (<http://www.idefics.eu>). We gratefully acknowledge the financial support of the European Community within the Sixth RTD Framework Programme Contract No. 016181 (FOOD). We are grateful to the Volkswagen Foundation that financially supported the production of this supplement. This analysis was also supported by a grant from the Spanish Ministry of Science and Innovation (RYC-2011-09011). LL received grant support from VR, FORTE and FORMAS (research councils in Sweden). The authors wish to thank the IDEFICS children and their parents who generously volunteered have participated in the IDEFICS study. We also acknowledge all the staff members involved in the fieldwork for their efforts and great enthusiasm. The information in this document reflects the authors' view and is provided as is.

STATEMENT OF ETHICS

We certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during this research. Approval by the appropriate Ethics Committees was obtained by each of the 8 centres doing the fieldwork. Study children did not undergo any procedure before both they and their parents had given consent for examinations, collection of samples, subsequent analysis and storage of personal data and collected samples. Study subjects and their parents could consent to single components of the study while abstaining from others.

REFERENCES

- 1 Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep* 1985; **100**: 126–131.
- 2 Ortega FB, Ruiz JR, Castillo MJ, Sjostrom M. Physical fitness in childhood and adolescence: a powerful marker of health. *Int J Obes (Lond)* 2008; **32**: 1–11.
- 3 Ekelund U, Anderssen SA, Froberg K, Sardinha LB, Andersen LB, Brage S. Independent associations of physical activity and cardiorespiratory fitness with metabolic risk factors in children: the European youth heart study. *Diabetologia* 2007; **50**: 1832–1840.
- 4 Blaes A, Baquet G, Fabre C, Van Praagh E, Berthoin S. Is there any relationship between physical activity level and patterns, and physical performance in children? *Int J Behav Nutr Phys Act* 2011; **8**: 122.
- 5 Ortega FB, Labayen I, Ruiz JR, Kurvinen E, Loit HM, Harro J *et al.* Improvements in fitness reduce the risk of becoming overweight across puberty. *Med Sci Sports Exerc* 2011; **43**: 1891–1897.
- 6 Ortega FB, Silventoinen K, Tynelius P, Rasmussen F. Muscular strength in male adolescents and premature death: cohort study of one million participants. *BMJ* 2012; **345**: e7279.
- 7 Moliner-Urdiales D, Ortega FB, Vicente-Rodriguez G, Rey-Lopez JP, Gracia-Marco L, Widhalm K *et al.* Association of physical activity with muscular strength and fat-free mass in adolescents: the HELENA study. *Eur J Appl Physiol* 2010; **109**: 1119–1127.
- 8 Morales PF, Sanchez-Lopez M, Moya-Martinez P, Garcia-Prieto JC, Martinez-Andres M, Garcia NL *et al.* Health-related quality of life, obesity, and fitness in schoolchildren: the Cuenca study. *Qual Life Res* 2013; **22**: 1515–1523.
- 9 Ruiz JR, Castro-Pinero J, Artero EG, Ortega FB, Sjostrom M, Suni J *et al.* Predictive validity of health-related fitness in youth: a systematic review. *Br J Sports Med* 2009; **43**: 909–923.

- 10 Catley MJ, Tomkinson GR. Normative health-related fitness values for children: analysis of 85347 test results on 9-17-year-old Australians since 1985. *Br J Sports Med* 2013; **47**: 98-108.
- 11 Olds T, Tomkinson G, Leger L, Cazorla G. Worldwide variation in the performance of children and adolescents: an analysis of 109 studies of the 20-m shuttle run test in 37 countries. *J Sports Sci* 2006; **24**: 1025-1038.
- 12 Pate RR, Wang CY, Dowda M, Farrell SW, O'Neill JR. Cardiorespiratory fitness levels among US youth 12 to 19 years of age: findings from the 1999-2002 National Health and Nutrition Examination Survey. *Arch Pediatr Adolesc Med* 2006; **160**: 1005-1012.
- 13 Carrel AL, Bowser J, White D, Moberg DP, Weaver B, Hisgen J et al. Standardized childhood fitness percentiles derived from school-based testing. *J Pediatr* 2012; **161**: 120-124.
- 14 Tremblay MS, Shields M, Laviolette M, Craig CL, Janssen I, Connor Gorber S. Fitness of Canadian children and youth: results from the 2007-2009 Canadian Health Measures Survey. *Health Rep* 2010; **21**: 7-20.
- 15 Ahrens W, Bammann K, Siani A, Buchecker K, De Henauw S, Iacoviello L et al. IDEFICS consortium. The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes (Lond)* 2011; **35**: S3-S15.
- 16 Ruiz JR, Castro-Pinero J, Espana-Romero V, Artero EG, Ortega FB, Cuenca MM et al. Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. *Br J Sports Med* 2011; **45**: 518-524.
- 17 Artero EG, Espana-Romero V, Castro-Pinero J, Ortega FB, Suni J, Castillo-Garzon MJ et al. Reliability of field-based fitness tests in youth. *Int J Sports Med* 2011; **32**: 159-169.
- 18 Castro-Pinero J, Artero EG, Espana-Romero V, Ortega FB, Sjostrom M, Suni J et al. Criterion-related validity of field-based fitness tests in youth: a systematic review. *Br J Sports Med* 2010; **44**: 934-943.
- 19 Leger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sci* 1988; **6**: 93-101.
- 20 Meredith, M Welk, G. *Fitnessgram-Activitygram Test Administration Manual*. Human Kinetics: Champaign, IL, USA, 2007.
- 21 Ruiz JR, Espana-Romero V, Ortega FB, Sjostrom M, Castillo MJ, Gutierrez A. Hand span influences optimal grip span in male and female teenagers. *J Hand Surg Am* 2006; **31**: 1367-1372.
- 22 Castro-Pinero J, Ortega FB, Artero EG, Girela-Rejon MJ, Mora J, Sjostrom M et al. Assessing muscular strength in youth: usefulness of standing long jump as a general index of muscular fitness. *J Strength Cond Res* 2010; **24**: 1810-1817.
- 23 Bammann K, Peplies J, Sjöström M, Lissner L, De Henauw S, Galli C et al. Assessment of diet, physical activity and biological, social and environmental factors in a multi-centre European project on diet- and lifestyle-related disorders in children (IDEFICS). *J Public Health* 2006; **14**: 279-289.
- 24 Stomfai S, Ahrens W, Bammann K, Kovacs E, Marild S, Michels N et al. IDEFICS consortium. Intra- and inter-observer reliability in anthropometric measurements in children. *Int J Obes (Lond)* 2011; **35**: S45-S51.
- 25 Stasinopoulos DM, Rigby R. Generalized additive models for location scale and shape (GAMLSS) in R. *J Stat Softw* 2007; **23**: 1-46.
- 26 R Core Team. R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2013. <http://www.R-project.org/>.
- 27 Van Buuren S, Fredriks A. Worm plot: a simple diagnostic device for modelling growth reference curves. *Stat Med* 2001; **20**: 1259-1277.
- 28 Cole TJ, Stanokevic S, Stocks J, Coates A, Hankinson J, Wade A. Age- and size-related reference ranges: A case study of spirometry through childhood and adulthood. *Stat Med* 2009; **28**: 880-898.
- 29 Ruiz JR, Ortega FB, Gutierrez A, Meusel D, Sjöström M, Castillo MJ. Health-related fitness assessment in childhood and adolescence: a European approach based on the AVENA, EYHS and HELENA studies. *J Public Health* 2006; **14**: 269-277.
- 30 Metter EJ, Talbot LA, Schrager M, Conwit R. Skeletal muscle strength as a predictor of all-cause mortality in healthy men. *J Gerontol A Biol Sci Med Sci* 2002; **57**: B359-B365.
- 31 Blair SN, Kohl 3rd HW, Paffenbarger Jr RS, Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* 1989; **262**: 2395-2401.
- 32 Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE. Exercise capacity and mortality among men referred for exercise testing. *N Engl J Med* 2002; **346**: 793-801.
- 33 Mora S, Redberg RF, Cui Y, Whiteman MK, Flaws JA, Sharrett AR et al. Ability of exercise testing to predict cardiovascular and all-cause death in asymptomatic women: a 20-year follow-up of the lipid research clinics prevalence study. *JAMA* 2003; **290**: 1600-1607.
- 34 Ara I, Moreno LA, Leiva MT, Gutin B, Casajus JA. Adiposity, physical activity, and physical fitness among children from Aragon, Spain. *Obesity (Silver Spring)* 2007; **15**: 1918-1924.
- 35 Nassis GP, Psarra G, Sidossis LS. Central and total adiposity are lower in overweight and obese children with high cardiorespiratory fitness. *Eur J Clin Nutr* 2005; **59**: 137-141.
- 36 Ruiz JR, Ortega FB, Rizzo NS, Villa I, Hurtig-Wennlof A, Oja L et al. High cardiovascular fitness is associated with low metabolic risk score in children: the European Youth Heart Study. *Pediatr Res* 2007; **61**: 350-355.
- 37 Artero EG, Ruiz JR, Ortega FB, Espana-Romero V, Vicente-Rodriguez G, Molnar D et al. Muscular and cardiorespiratory fitness are independently associated with metabolic risk in adolescents: the HELENA study. *Pediatr Diabetes* 2011; **12**: 704-712.
- 38 Vicente-Rodriguez G, Urzanqui A, Mesana MI, Ortega FB, Ruiz JR, Ezquerro J et al. Physical fitness effect on bone mass is mediated by the independent association between lean mass and bone mass through adolescence: a cross-sectional study. *J Bone Miner Metab* 2008; **26**: 288-294.
- 39 Castelli DM, Hillman CH, Buck SM, Erwin HE. Physical fitness and academic achievement in third- and fifth-grade students. *J Sport Exerc Psychol* 2007; **29**: 239-252.
- 40 Moliner-Urdiales D, Ruiz JR, Vicente-Rodriguez G, Ortega FB, Rey-Lopez JP, Espana-Romero V et al. Associations of muscular and cardiorespiratory fitness with total and central body fat in adolescents: the HELENA study. *Br J Sports Med* 2011; **45**: 101-108.
- 41 Twisk JW, Kemper HC, van Mechelen W. Tracking of activity and fitness and the relationship with cardiovascular disease risk factors. *Med Sci Sports Exerc* 2000; **32**: 1455-1461.
- 42 Casajus JA, Ortega FB, Vicente-Rodriguez G, Leiva MT, Moreno LA, Ara I. [Physical fitness, fat distribution and health in school-age children (7 to 12 years).] *Rev Int Med Cienc Act Fis Deporte* 2010; **12**: 523-537.
- 43 Castro-Pinero J, Gonzalez-Montesinos JL, Mora J, Keating XD, Girela-Rejon MJ, Sjostrom M et al. Percentile values for muscular strength field tests in children aged 6 to 17 years: influence of weight status. *J Strength Cond Res* 2009; **23**: 2295-2310.
- 44 Castro-Pineiro J, Ortega FB, Keating XD, Gonzalez-Montesinos JL, Sjostrom M, Ruiz JR. Percentile values for aerobic performance running/walking field tests in children aged 6 to 17 years: influence of weight status. *Nutr Hosp* 2011; **26**: 572-578.
- 45 Ortega FB, Artero EG, Ruiz JR, Espana-Romero V, Jimenez-Pavon D, Vicente-Rodriguez G et al. Physical fitness levels among European adolescents: the HELENA study. *Br J Sports Med* 2011; **45**: 20-29.
- 46 Haugen T, Hoigaard R, Seiler S. Normative data of BMI and physical fitness in a Norwegian sample of early adolescents. *Scand J Public Health* 2014; **42**: 67-73.
- 47 Hasselstrom H, Hansen SE, Froberg K, Andersen LB. Physical fitness and physical activity during adolescence as predictors of cardiovascular disease risk in young adulthood. Danish Youth and Sports Study. An eight-year follow-up study. *Int J Sports Med* 2002; **23**: S27-S31.
- 48 Twisk JW, Kemper HC, van Mechelen W. Prediction of cardiovascular disease risk factors later in life by physical activity and physical fitness in youth: general comments and conclusions. *Int J Sports Med* 2002; **23**: S44-S49.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Throwing and Kicking Ability of 8-Year-Old Boys and Girls¹

PAUL DOHRMANN

Illinois State University
Normal, Illinois

Abstract

Two hundred 8-year-old subjects were placed into two groups, each containing 50 boys and 50 girls. Both groups participated in their regular physical education program throughout the entire 36 weeks of the study. The fall training group engaged in a training program for throwing and kicking for distance during the first 9 weeks of the study. The spring training group engaged in the same training program during the last 9 weeks. The findings indicate that throwing and kicking training programs, in addition to regular physical education programs, do not result in greater improvements in throwing and kicking ability than do regular physical education programs; and that in throwing and in kicking ability 8-year-old boys are superior to 8-year-old girls ($P = .01$).

SEVERAL STUDIES CONCERNING CHILDREN aged from three to eight years (2, 3, 5), indicate that boys are superior to girls in throwing for distance and kicking for distance. Dusenberry (1) found that training in throwing for distance results in gains above the effects of maturation and general practice for children aged three to seven. Questions might be raised on whether similar results would be obtained for 8-year-old children who are participating in a daily program of physical education and on when during the eighth year such a program would be most effective.

The purpose of this study was to investigate the throwing and kicking ability of 8-year-old boys and 8-year-old girls and to determine the effects that training programs in throwing and kicking have upon the ability to throw and to kick for distance.

Procedure

All subjects were enrolled in the elementary schools of University City, Missouri. They were in good health and had participated in the regular physical education program for the first and second grades prior to the study. The subjects were considered to be 8 years old if at the beginning of the study their ages were between 7 years 9 months and 8 years 3 months. Fifty boys and 50 girls were placed in the fall training group. Every effort was

¹ This is part of a report of a study submitted in partial fulfillment of the requirements for the Ph.D. degree in the Division of Physical Education, Graduate College, State University of Iowa, Iowa City, 1961.

Throwing and Kicking Ability

465

made to prevent the spring training group from knowing what the fall training group was doing.

Both groups were tested for throwing and for kicking for distance at the beginning of the study and at the end of the first 9 weeks. Immediately after the initial testing, the fall training group (T_1) participated in a 9-week training program in addition to their regular physical education program. Once each 3 weeks throughout the training program the fall training group was retested with the same tests as were used in the initial testing.

At the end of the first 9 weeks, both groups continued to participate in the regular physical education program until the spring of the following year, at which time the procedures for the groups were reversed. During the last 9 weeks of the study, the spring training group (T_2) participated in the same training program that the fall training-group (T_1) had participated in during the first 9 weeks. Both groups were tested in the same manner as in the fall.

TESTS

Two tests were used in the testing program: a throw for distance using a 12-in. softball and a kick for distance using a soccer ball. For each test the best score out of 10 trials was recorded. The tests were administered outdoors, on blacktop surfaces marked at 3-ft. intervals with chalk lines. Each throw was scored as the distance between the front edge of the forward foot and the spot where the ball landed. In the kick for distance, the ball was placed upon a platform 1-in. high, which was formed on the first chalk line by placing two felt blackboard erasers (3 in. x 6 in. x 1 in.) adjacent and parallel to each other. The ball was elevated in order to aid the subject in raising the ball off the ground.

Estimates of the reliability of the throwing for distance and the kicking for distance tests, for both boys and girls, were obtained by the test-retest technique. An r of .95 for throwing and of .72 for kicking was obtained for the boys and an r of .93 for throwing and of .71 for kicking was obtained for the girls.

TRAINING PROCEDURE

The training program consisted of throwing for distance and kicking for distance on alternate days for 9 weeks. In each training period for throwing, the subjects made 20 throws for distance and were coached in the mechanics of throwing. In each training period for kicking, the subjects made 20 kicks for distance and were coached in the mechanics of kicking. An attempt was made to motivate the subjects to do their very best in each attempt.

The training program was administered by the regular physical education teacher. Before the experiment the teachers were given mimeographed instructions which included information about the experiment, a detailed plan for conducting the training program, and a list of the coaching points in the mechanics of kicking and throwing which were to be stressed.

The nontraining group participated in the regular physical education pro-

TABLE 1.—THROWING AND KICKING SCORES

Test	Sex	N	Mean (in.)	Minimum	25	Percentiles 50	75	Maximum
Initial fall throwing	Boys	100	672	333	524	640	809	1,195
	Girls	100	348	105	283	333	398	687
Final spring throwing	Boys	100	805	369	650	776	955	1,340
	Girls	100	441	238	353	414	487	925
Initial fall kicking	Boys	100	333	90	233	301	425	670
	Girls	100	230	53	169	222	284	601
Final spring kicking	Boys	100	404	144	314	403	474	852
	Girls	100	293	108	203	276	360	660

TABLE 2.—INITIAL MEANS, FINAL MEANS, AND ADJUSTED MEANS FOR
THROWING AND KICKING SCORES

Test	Group	N	M _x (in.)	M _y (in.)	M' _y (in.)
Final fall throwing	T ₁	100	488	498	518
	T ₂	100	531	524	504
Initial spring throwing	T ₁	100	488	537	556
	T ₂	100	531	592	573
Final spring throwing	T ₁	100	488	586	608
	T ₂	100	531	659	638
Final fall kicking	T ₁	100	250	311	331
	T ₂	100	313	330	310
Initial spring kicking	T ₁	100	250	335	356
	T ₂	100	313	365	344
Final spring kicking	T ₁	100	250	315	334
	T ₂	100	313	387	366

Throwing and Kicking Ability

467

gram for the 9-week period. This program consisted of a daily 20-min. period of instruction supervised by the physical education teacher, and a 15-min. play period supervised by the classroom teacher.

Analysis

Analysis of variance, according to the method described by Lindquist (4), was employed to determine if there were any differences between the boys and the girls in their ability to throw and kick for distance. In each analysis the F for the difference between the means of the scores for the boys and for the girls was computed. An F of 6.85 ($df = 1,120$) corresponding to the .01 level of significance was chosen to designate the region of rejection for the null hypothesis. The means, minimum scores, maximum scores, and selected percentiles for the initial fall and for the final spring tests are presented in Table 1.

Analysis of covariance, according to the method described by Lindquist (4), using the initial fall scores as the control variable was used to deter-

TABLE 3.—PERCENTAGE OF BOYS AND OF GIRLS EXECUTING BEST THROW AND BEST KICK AND CUMULATIVE PERCENTAGE

Test	Boys		Girls	
	Percent for Test	Cumulative Percent	Percent for Test	Cumulative Percent
Fall training program throwing				
Initial	24	24	18	18
At end of 3 weeks	30	54	34	52
At end of 6 weeks	22	76	18	70
Final	24	100	30	100
Spring training program throwing				
Initial	10	10	4	4
At end of 3 weeks	10	20	12	16
At end of 6 weeks	26	46	36	52
Final	54	100	48	100
Fall training program kicking				
Initial	22	22	12	12
At end of 3 weeks	12	34	22	34
At end of 6 weeks	10	44	26	60
Final	56	100	40	100
Spring training program kicking				
Initial	28	28	8	8
At end of 3 weeks	22	50	22	30
At end of 6 weeks	20	70	32	62
Final	30	100	38	100

TABLE 4.—SUMMARY OF ANALYSIS OF VARIANCE OF THROWING SCORES

Source	Initial Fall			Final Fall			Initial Spring			Final Spring		
	df	ms	F	df	ms	F	df	ms	F	df	ms	F
Treatment	1			1			1			1		
Sex	1	5,239,408	225.69*	1	5,444,340	207.80*	1	6,442,255	250.63*	1	6,627,712	214.05*
T × S	1			1			1			1		
Within	196	23,725		196	26,200		196	25,704		196	30,963	
Total	199			199			199			199		

* P = .001.

TABLE 5.—SUMMARY OF ANALYSIS OF VARIANCE OF KICKING SCORES

Source	Initial Fall			Final Fall			Initial Spring			Final Spring		
	df	ms	F	df	ms	F	df	ms	F	df	ms	F
Treatment	1			1			1			1		
Sex	1	536,337	46.93*	1	688,433	58.96*	1	757,065	54.01*	1	666,666	55.64*
T × S	1			1			1			1		
Within	196	11,428		196	11,677		196	14,017		196	11,981	
Total	199			199			199			199		

* P = .001.

Throwing and Kicking Ability

469

TABLE 6.—SUMMARY OF ANALYSIS OF COVARIANCE OF THROWING SCORES

Source	Final Fall			Initial Spring			Final Spring		
	df	ms'	F	df	ms'	F	df	ms'	F
Treatment	1	10,376	2.2427	1	14,942	1.9590	1	46,950	5.5513
Sex	1			1			1		
T × S	1	2,336	.5049	1	21,369	2.8028	1	5,189	.6136
Within	195	4,626		195	7,627		195	8,457	
Total	198			198			198		

TABLE 7.—SUMMARY OF ANALYSIS OF COVARIANCE OF KICKING SCORES

Source	Final Fall			Initial Spring			Final Spring		
	df	ms'	F	df	ms'	F	df	ms'	F
Treatment	1	20,322	2.8835	1	6,128	.6849	1	47,785	6.3950
Sex	1			1			1		
T × S	1	13,757	1.9520	1	49,782	5.5641	1	5,212	.6976
Within	195	7,047		195	8,947		195	7,472	
Total	198			198			198		

mine if there were any differences between the fall training group (T_1) and the spring training group (T_2) in the ability to throw for distance and to kick for distance. For each analysis the statistical significance of the treatment effect, and that of the interaction effect (TS), was tested by the use of a mean-square ratio (F). An F of 6.85 ($df = 1, 120$), corresponding to the .01 level of significance, was chosen to designate the region of rejection for the null hypothesis. This small value (i.e., $P = .01$) was chosen because the same data were used in four analyses. The final means (M_y) were adjusted (M'_y) to allow for differences in the initial means (M_x) (Table 2).

For each of the four tests (i.e., initial, at end of 3 weeks, at end of 6 weeks, and final) given during the training program for throwing and for kicking, the percentages were computed for boys, and for those girls who during the test executed their best throws or kicks, together with the cumulative percentages (Table 3).

Results and Discussion

In throwing and in kicking ability 8-year-old boys were found to be superior ($P = .01$) to 8-year-old girls (Tables 4 and 5).

No difference was found between the improvement in throwing ability made during the first 9 weeks of the experiment by the fall training group (T_1), which received training for throwing in addition to that in the regular physical education program, and that made by the spring training group (T_2), which during the 9-week period received no training in addition to that in the regular physical education program (Table 6).

No difference was found between the improvement in throwing ability made during the last 9 weeks of the experiment by the spring training group (T_2), which received training for throwing in addition to that in the regular program of physical education, and that made by the fall training group (T_1), which during the 9-week period received no training in addition to that in the regular physical education program (Table 6).

For each of the three analyses of kicking scores, no difference was found between the mean of the kicking scores for T_1 , and the mean of the kicking scores for T_2 (Table 7). The implication for this finding is that, on the basis of kicking tests administered in the fall and in the spring, the kicking ability of 8-year-old children who participated in the fall training program (T_1) in addition to their regular physical education program, does not differ from that of 8-year-old children who have participated in the spring training program (T_2), in addition to their regular physical education program.

During the fall training program most of the boys and most of the girls completed their best throws by the end of 6 weeks (Table 3); during the spring training program, about half of the boys and half of the girls executed their best throws in the final tests (Table 3). This might imply that

if such a training program was utilized for children in the early part of their eighth year, it should be limited to 6 weeks in length, but if it was utilized for children late in their eighth year, the program should be continued for at least 9 weeks.

In a comparison of the two training programs for kicking for distance it is interesting to note that the percentages of boys and of girls who had completed their best kicks by the end of 6 weeks are slightly higher for the spring training program than for the fall training program (Table 3). This finding might imply that if a training program for kicking were utilized for children in the eighth year, it should come late in the year and be limited to 6 weeks.

The initial throwing and kicking scores of the boys are superior to those of the girls. For this reason the girls might have been expected to improve more than the boys. However, changes in throwing and kicking scores during the training program for throwing and kicking for distance are the same for the girls as for the boys. This finding is indicated by the absence of a statistically significant interaction effect (Tables 6 and 7).

Conclusions

The findings in this study would appear to warrant the conclusions that for 8-year-old boys and girls throwing and kicking training programs, in addition to regular physical education programs, do not result in greater improvements in throwing and in kicking ability than do regular physical education programs; and that in throwing and in kicking ability 8-year-old boys are superior to 8-year-old girls ($P = .01$).

References

1. DUSENBERRY, LOIS. A study of the effects of training in ball throwing by children ages three to seven. *Res. Quart.* 23:9-14, 1952.
2. HARTMAN, DORIS. The hurdle jump as a measure of motor efficiency of young children. *Child Develpm.* 14:90-105, 1943.
3. JENKINS, LULU. *A comparative study of motor achievements of children five, six, and seven years of age.* New York City: Bureau of Publications, Teachers College, Columbia University, 1930.
4. LINDQUIST, E. F. *Design and analysis of experiments in psychology and Education.* Boston: Houghton Mifflin Co., 1953.
5. SEILS, LEROY. The relationship between measure of physical growth and gross motor performance of primary-grade school children. *Res. Quart.* 22:244-60, 1951.

(Submitted December 18, 1963)

ORIGINAL ARTICLE

Maximum oxygen uptake and objectively measured physical activity in Danish children 6–7 years of age: the Copenhagen school child intervention study

S Eiberg, H Hasselstrom, V Grønfeldt, K Froberg, J Svensson, L B Andersen

Br J Sports Med 2005;**39**:725–730. doi: 10.1136/bjsm.2004.015230

See end of article for authors' affiliations

Correspondence to: Professor Andersen, Department of Health, Norwegian University of Sport and Physical Education, Oslo 0806, Norway; lars.bo.andersen@nih.no

Accepted 29 March 2005

Objectives: To provide normative data on maximum oxygen uptake ($\text{VO}_{2\text{MAX}}$) and physical activity in children 6–7 years of age and analyse the association between these variables.

Methods: $\text{VO}_{2\text{MAX}}$ was measured in 366 boys (mean (SD) 6.8 (0.4) years of age) and 332 girls (6.7 (0.4) years of age) from preschool classes in two suburban communities in Copenhagen, during a progressive treadmill exercise. Habitual physical activity was measured with accelerometers.

Results: Boys had higher $\text{VO}_{2\text{MAX}}$ both in absolute values (1.19 (0.18) v 1.06 (0.16) litres/min (+11%), $p < 0.001$) and relative to body weight (48.5 (6.0) v 44.8 (5.6) ml/kg/min (+8%); $p < 0.001$) than girls. The difference in $\text{VO}_{2\text{MAX}}$ between boys and girls decreased to +2% when expressed relative to lean body mass (LBM). Absolute $\text{VO}_{2\text{MAX}}$ was related to LBM, body mass, and stature (all $p < 0.001$). Boys were more physically active than girls (mean counts +9.4%, $p < 0.001$), and even when boys and girls with the same $\text{VO}_{2\text{MAX}}$ were compared, boys were more active. The difference in physical activity between the sexes was higher when sustained activity of higher intensity was compared.

Conclusions: $\text{VO}_{2\text{MAX}}$ is higher in boys than girls (+11%), even when related to body mass (+8%) and LBM (+2%). Most of the difference in $\text{VO}_{2\text{MAX}}$ relative to body mass was explained by the larger percentage body fat in girls. When boys and girls with the same $\text{VO}_{2\text{MAX}}$ were compared, boys engaged in more minutes of exercise of at least moderate intensity.

The fitness level in children in the western world has declined, but few data are available.¹ Even though children do not suffer from lifestyle diseases, it is recommended that they stay physically active and fit, because a high physical activity (PA) level may prevent future illness.² A high level of physical fitness has been associated with a decreased risk of cardiovascular disease (CVD) in the adult population.³ Even in children, risk factors for CVD have been identified,⁴ and physical fitness also seems to have an effect on the level of risk factors in children.⁵ Even though children do not suffer from CVD, it is recommended that children are physically active and fit, as this may help to prevent the disease in the future. Therefore it is important to assess physical fitness at an early age. Maximum oxygen uptake ($\text{VO}_{2\text{MAX}}$) is probably the best index of physical fitness and has been studied intensively in adults.⁶ However, measurements of $\text{VO}_{2\text{MAX}}$ in 6–7 year old children are sparse, and representative data do not exist.^{6–11} Even in older children and adolescents, very few population based studies exist.^{10–12–13} The reasons for this include difficulties in testing young children. Ethical considerations, safety factors, and equipment made for adults render testing with young children more challenging. Another health related measure of children is their level of PA, as PA has a positive effect on metabolism.¹⁴ Most studies have assessed PA by self report, but accelerometers provide a robust measure of habitual PA.¹⁵ No population study has analysed the association between $\text{VO}_{2\text{MAX}}$ and PA assessed by accelerometry.

$\text{VO}_{2\text{MAX}}$ is consistently higher in boys than in girls even before puberty.¹⁶ This has been attributed to a different body composition and a larger stroke volume of the heart in boys. However, even though boys have been shown to be more active than girls, the difference is too small to explain the difference in $\text{VO}_{2\text{MAX}}$.^{8–17–18}

The aim of this study was to provide population data on $\text{VO}_{2\text{MAX}}$ and PA measured using accelerometry in 6–7 year old children from Copenhagen, and also explore potential sex differences and analyse the association between PA and $\text{VO}_{2\text{MAX}}$. We hypothesised that boys have higher $\text{VO}_{2\text{MAX}}$ than girls, and that this difference may be the result of differences in their levels of fat and PA.

METHODS

Children from 46 preschool classes (6–7 years of age) in 18 schools in two suburban communities in the Copenhagen area were invited to participate in the Copenhagen school child intervention study. In 2000 the community of Ballerup (10 schools, 27 classes) increased the number of PE lessons from two to four a week for the first three years of school. The community of Tårnby (eight schools, 19 classes) was chosen as a control as it resembles Ballerup in sociodemographics. A total of 706 children (69% of those eligible) volunteered for the study, and written informed consent was obtained from the parents/guardians. Of these 706 children, 415 from Ballerup and 291 from Tårnby participated. The ethics committee of Copenhagen county approved the study. The tests were performed from December 2001 until May 2002 at the 18 different schools involved. The exercise test was performed using permanently installed equipment in a camper trailer. All other physiological tests were performed in a gym or a classroom. All tests were performed before noon (0800–1200).

PA was generally measured one week after the other tests. At schools with more than 50 participating pupils, two weeks of PA measurements were necessary. The 31% not

Abbreviations: BMI, body mass index; CVD, cardiovascular disease; LBM, lean body mass; PA, physical activity; SFS, sum of four skinfolds; $\text{VO}_{2\text{MAX}}$, maximum oxygen uptake

participating were analysed about one year later. The first systematic medical examination of children in Denmark takes place one and a half years after they start school, so we were unable to gather data from non-participants in their first year. There were no significant differences between groups with respect to age, height, weight, and body mass index (BMI) for either sex. The analysis included 612 participants and 277 non-participants.

Height was measured by a Harpenden stadiometer to the nearest 1 mm. Body weight was measured to the nearest 0.1 kg using a SECA electronic scale. Bicipital, tricipital, subscapular, and suprailiac skinfolds were measured with a Harpenden skinfold caliper according to criteria presented by De Lorenzo *et al.*¹⁹ The dominant side of the body was determined by asking the child to take a pen and write his/her name. The data shown in this study represent the mean of three measurements taken on the non-dominant side of the body. The sum of four skinfolds (SFS) was used as an indicator of body fatness. Fat mass, fat percentage, and lean body mass (LBM) were derived as:

Fat mass (kg) = $0.38 \times \text{body weight} + (0.30 \times \text{triceps}) + (0.87 \times G) - 9.42$

where for boys $G = 1$ and for girls $G = 2$.²⁰

Habitual PA was measured by the MTI 7164 activity monitor (Manufactory Technology Inc, Fort Walton Beach, Florida, USA). The monitor has been validated in several studies and has shown both high mechanical reproducibility¹⁵ and good validity with respect to free living conditions in children.²¹ The monitor samples acceleration at 10 Hz and integrates counts over a time period (epoch) defined by the user. In children, PA is characterised by short bursts of activity. Therefore we chose an epoch of 10 seconds. The choice of epoch limited the registration to three full days and 19 hours. To allow familiarisation, the children had the MTI monitor put on one day before recording. It was secured directly to the skin at the lower back using an elastic belt. The children were instructed to wear the monitor continuously except during water based activity or when sleeping. To distinguish true zeros that arise as a result of below threshold activity from zeros recorded when the MTI monitor was not worn, the data were cleaned as follows: all MTI files were screened for sustained periods of zero activity. Periods of 10 minutes or more with zero counts were interpreted as "MTI not worn" and removed from the file. In spite of instructions, some children slept with the monitor on and therefore had activity recorded even late at night. Therefore we chose to control for nocturnal activity (2230 to 0600). Given these criteria, the data were included if the child had accumulated more than eight hours of activity a day for at least three days. In the end, 466 children had four valid days and 96 children had three valid days, of whom 82 had two week days and one weekend day and 14 had one week day and two weekend days. Fifty eight children had less than three valid days. A mean count was calculated for each child. Furthermore, the number of minutes in periods longer than five or 10 minutes of sustained activity above 2000 counts/min, which is about three METS, was calculated.

Measurement of $\dot{V}O_{2\text{MAX}}$

$\dot{V}O_2$ was determined with an AMIS 2001 Cardiopulmonary Function Test System (Innovision, DK 5260 Odense, Denmark). This system has been validated against the Douglas bag system.²² We used a Hans Rudolph mouth piece with a volume of 15 ml especially designed for children. Heart rate was measured continuously every fifth second (Polar Sport Tester, Kempele, Finland).

We conducted a pilot study to establish a protocol that enabled most of the children to reach exhaustion. We chose a continuous walking and running protocol on a treadmill. The

velocity on the treadmill was initially set to 4 km/h without inclination and kept there for the first three minutes to allow familiarisation. At three minutes, the velocity was increased to 8 km/h, and at five minutes the inclination was raised to 3%. At seven, nine, and 11 minutes, the inclination was increased to 6%, 9%, and 11% respectively. If the child could endure more, the velocity was increased to 9 km/h after 13 minutes and then 10 km/h at 15 minutes. No child completed the last work load.

The children were instructed to run until exhaustion. One subjective and three objective criteria were used to determine if the test was maximal. Every child had to meet the subjective criterion and at least one of the three objective criteria. Criteria were chosen according to Rowland's recommendations.²³ The physiological criteria were: heart rate >200 beats/min; respiratory exchange ratio ≥ 0.99 ; a defined plateau of $\dot{V}O_2$ (an increase less than 2.1 ml/kg/min). Subjective criteria were signs of intense effort such as unsteady running pattern, sweating, facial flushing, and clear unwillingness to continue running in spite of repeated strong verbal encouragement.

Twenty 6–7 year old children outside the sample were invited to do a test-retest on the peak $\dot{V}O_2$ test. Written informed consent was obtained from the parents/guardians. The children were retested either five or seven days after the first test. Four children did not meet the criteria on one or both occasions, and their results were not used in the later analysis, leaving 16 children with two valid tests. A Bland-Altman plot showed no systematic difference between the first and second test, suggesting no learning effect from the first to the second test. The plot also showed that the size of the difference was not dependent on the absolute level of $\dot{V}O_{2\text{MAX}}$. The typical error of measurement on the difference between $\dot{V}O_{2\text{MAX}}$ between tests was 0.02 litre/kg.

Statistical analysis

The data were stored and analysed using SPSS 11.5.0. Data that were not normally distributed were log transformed, and the mean of transformed values was back transformed to obtain the geometric mean as suggested by Altman.²⁴ $\dot{V}O_{2\text{MAX}}$ relative to body mass was split into deciles by sex to compare differences over the whole distribution, and $\dot{V}O_{2\text{MAX}}$ (ml/kg/min) was calculated for each of the deciles. For a comparison of differences between sexes, we adjusted for body size by two methods of allometric scaling. Data from this study were used to find the scaling factors. In the analyses, the scaling exponent b was identified in the allometric equation $Y = a_1 + a_2 X^b$, where Y is the physiological variable ($\dot{V}O_{2\text{MAX}}$) and X is the anthropometric scaling variable (weight or LBM). To obtain b , both the Y and X were log transformed, and least squares regression identified the b in the equation $\ln(Y) = \ln(a_2) + b \ln(X)$.

Differences between sexes were tested using Student's t test. The relation between $\dot{V}O_{2\text{MAX}}$ and PA was assessed by linear regression. A significance level of $p < 0.05$ was chosen.

RESULTS

Table 1 shows the total number of boys and girls that entered the study, the number of valid tests, the number of children that stopped before exhaustion, the number of children not fulfilling the criteria, and children who refused to do the treadmill test. A valid $\dot{V}O_{2\text{MAX}}$ measurement was not assessed in 114 children, because 37 failed to comply with the approval criteria and 57 did not want to wear the $\dot{V}O_2$ equipment or did not wear the equipment satisfactorily. Ten children were unwilling to perform the test, and 10 children were absent on all the test days. The final number of accepted tests was 592. Analysis of children with non-valid tests was performed for each sex separately. Time to exhaustion, heart

Table 1 Age, height, body weight, and maximal exercise data for groups with different test status for the maximal exercise test

	Boys				Girls			
	Valid	Non-valid	No VO_2 data†	Not willing to be tested	Valid	Non-valid	No VO_2 data†	Not willing to be tested
Number	309	20	28	7	283	17	29	3
Age (years)	6.8 (0.4)	6.7 (0.3)	6.7 (0.4)	6.9 (0.4)	6.7 (0.4)	6.6 (0.3)	6.6 (0.4)	6.4 (0.4)
Stature (m)	1.24 (0.05)	1.23 (0.05)	1.23 (0.06)	1.23 (0.05)	1.22 (0.05)	1.20 (0.04)*	1.23 (0.04)	1.21 (0.01)
Mass (kg)‡	24.6 (3.7)	24.4 (3.3)	24.7 (3.2)	24.3 (4.6)	23.9 (3.2)	22.6 (3.2)	25.0 (4.5)	22.8 (1.3)
Heart rate (beats/min)	196 (9)	174 (8) *	193 (9)		199 (9)	179 (6) *	196 (11)	
Test time (min:s)	8:06 (1:55)	5:53 (1:16)*	7:39 (1:39)		7:28 (1:36)	4:50 (0:55)*	6:46 (1:33)*	
VO_2MAX (litres/min)	1.19 (0.18)	0.99 (0.24)*			1.06 (0.16)	0.76 (0.16)*		

Values are mean (SD).

*Significant difference between valid test groups and other groups: $p < 0.05$.

†Children performing a maximal test with no recording of VO_2 .

‡Mass was not a Gaussian distribution. Therefore data were log transformed and a mean was calculated and then back transformed to a geometric mean.

rate, and VO_2MAX were lower for the 20 boys and 17 girls that did not meet the criteria ($p < 0.001$). The girls that did not comply with the criteria were 2.6 cm shorter than girls who did comply ($p < 0.05$). Twenty eight boys and 29 girls had a recorded time to exhaustion, but had no valid VO_2 measurements. The girls without a valid VO_2 recording ran 42 seconds less ($p < 0.05$) than girls with a valid test. Differences between all other age, anthropometric, and VO_2 test data were non-significant. Only data for the 592 children with a valid test were included in the following analysis. Of the 592 children that met the subjective criteria, 210 met the levelling off criterion (35%), 283 (47%) met the pulse criterion, and 507 (86%) met the respiratory exchange ratio criterion.

Table 2 shows descriptive data for age, anthropometry, and PA. Except for BMI, differences were found between boys and girls for all variables. The girls had larger SFS (15%, $p < 0.001$) and fat percentage (23%, $p < 0.001$). The boys were older (2%, $p < 0.001$), heavier (3%, $p < 0.05$), taller (1%, $p < 0.001$), had a greater LBM (8%, $p < 0.001$), a higher level of PA (mean counts/min 8%, $p < 0.001$), and more minutes in periods of five or 10 minutes of sustained activity at a level of 2000 counts/min or greater (60% and 103% respectively, $p < 0.001$). It should be noted that all boys had five minute periods of sustained activity, whereas three girls did not have any. Further, 38 boys and 83 girls did not have any 10 minute periods of sustained activity.

The absolute differences in the treadmill data between boys and girls were moderate (table 3). The girls had higher maximum heart rate (2%, $p < 0.001$) and respiratory exchange ratio (5%, $p < 0.001$), and the boys had higher absolute VO_2MAX (11%, $p < 0.001$), VO_2MAX relative to body mass (8%, $p < 0.001$), and VO_2MAX relative to LBM (2%, $p < 0.05$). VO_2MAX remained different (9%, $p < 0.001$) after

allometric scaling using (body mass)^{0.712} as scaling, but did not differ when LBM^{1.105} was used as the scaling factor.

Deciles in VO_2MAX (ml/kg/min) by sex were constructed and then VO_2MAX was compared between boys and girls for each decile (fig 1). This was done to see if the difference between the sexes appeared over the whole distribution or if only the lower or upper part of the distribution differed. The whole distribution was shifted to the right in boys compared with girls. Differences were found between all groups within sexes using Bonferroni post hoc analysis of variance ($p < 0.001$).

Relation between PA and fitness

A scatterplot of VO_2MAX relative to body weight and the PA variables showed a linear relation. A linear regression was performed with mean count of PA, the number of minutes above 2000 counts/min in five minute periods, and sex as independent variables, and VO_2MAX relative to body weight as a dependent variable.

The following equations where $G = 1$ for boys and $G = 2$ for girls were used:

Model (1) VO_2MAX (l/kg/min) = $49.9 - 3.8 \times G + 0.02 \times \text{PA}_{\text{mean count}}$ ($p < 0.001$)

Model (2) VO_2MAX (l/kg/min) = $49.6 - 3.1 \times G + 0.02 \times \text{PA}_{\text{in 5 min periods}}$ ($p < 0.001$).

Partial correlations for model (1) were -0.38 and 0.12 for sex and $\text{PA}_{\text{mean count}}$ respectively. Partial correlations for model (2) were -0.25 and 0.24 for sex and $\text{PA}_{\text{in 5 min periods}}$ respectively.

To explore the relation VO_2MAX versus PA and fatness further, all subjects were ranked into six fitness groups by VO_2MAX (independent of sex) and the mean sums of PA level ($\text{PA}_{\text{in 5 min periods}}$) and SFS were calculated for each sex

Table 2 Anthropometric and physical activity data in 6–7 year old children by sex

	Boys		Girls		p Value
	No	Mean	No	Mean	
Age (years)	309	6.8 (0.4)	283	6.7 (0.4)	<0.05
Mass (kg)	309	24.4 (19.8 to 34.9)	283	23.7 (18.4 to 32.0)	<0.001
Stature (m)	309	1.24 (0.05)	283	1.22 (0.05)	<0.001
BMI (kg/m^2)	309	15.9 (13.5 to 20.2)	283	15.9 (13.2 to 20.5)	NS
SFS (mm)	307	23.3 (15.3 to 51.2)	281	27.5 (17.6 to 54.4)	<0.001
Fat percentage	307	16.5 (5.2)	281	21.4 (5.0)	<0.001
LBM (kg)	307	20.4 (1.8)	281	18.7 (1.9)	<0.001
PA (mean counts/min)	291	743 (452 to 1308)	269	679 (397 to 1062)	<0.001
PA (minutes in 5 min bouts)	291	122 (17 to 334)	269	76 (0 to 194)	<0.001
PA (minutes in 10 min bouts)	291	53 (0 to 194)	269	26 (0 to 94)	<0.001

Values in parentheses are SD or 95% confidence interval. Mass, BMI, SFS, and PA were not normally distributed. Therefore data were log transformed and a mean was calculated and then back transformed to a geometric mean as described in the text. For these variables a 95% confidence interval is given.

BMI, Body mass index; SFS, sum of four skinfolds; LBM, lean body mass; PA, physical activity.

Table 3 Maximal exercise data on 6–7 year old children by sex

	Boys (n = 309)	Girls (n = 283)	p Value
Heart rate (beats/min)	196 (9)	199 (9)	<0.001
RER	1.05 (0.11)	1.10 (0.11)	<0.001
VO ₂ MAX (litres/min)	1.19 (0.18)	1.06 (0.16)	<0.001
VO ₂ MAX relative to body mass (ml/kg/min)	48.5 (6.0)	44.8 (5.6)	<0.001
VO ₂ MAX relative to LBM (ml/kg/min)	58.2 (6.8)	57.0 (6.5)	<0.05
VO ₂ MAX relative to body mass (ml/kg ^{0.71} /min)	121.7 (14.5)	111.2 (12.9)	<0.001
VO ₂ MAX relative to LBM (ml/kg ^{1.105} /min)	42.4 (4.9)	41.9 (4.8)	NS

Values are mean (SD).

RER, Respiratory exchange ratio.

within each fitness group. When PA was compared between boys and girls with the same level of fitness, boys were more active than girls except for the least fit decile (fig 2). In a linear regression, the association between sex and PA level was highly significant ($r = 0.20$, $p < 0.001$) even after adjustment for fitness ($r = 0.26$, $p < 0.001$).

When fitness was compared between boys and girls with the same level of fitness, girls had a larger SFS than boys except for the least fit decile (fig 3). In a linear regression, the association between sex and SFS was borderline significant ($r = 0.09$, $p = 0.05$) after adjustment for fitness ($r = 0.27$, $p < 0.001$). After adjustment for both fitness and PA, there was no difference between sexes with respect to SFS.

There were only two girls in the most fit group, and no standard error was plotted for this group in fig 3. The opposite analysis was also performed—that is, mean VO₂MAX was calculated for boys and girls with the same activity level and the same SFS respectively. The boys had a 3–5 ml/kg/min higher fitness level in each category of activity ($p < 0.001$) and a 1–6 ml/kg/min higher fitness in each category of SFS ($p < 0.01$).

DISCUSSION

The main findings in this study are that, in a large sample of young children, boys had a higher VO₂MAX and PA level than girls. Most of the difference in fitness between the sexes could be accounted for partly by body composition and partly by PA. As no differences in haemoglobin or sex hormones have been reported in this age group,¹⁷ it is likely that the difference in VO₂MAX relative to body weight is due to body composition and PA. However, when children with the same VO₂MAX were compared, boys were still more active, and in boys and girls with the same PA level, boys were fitter.

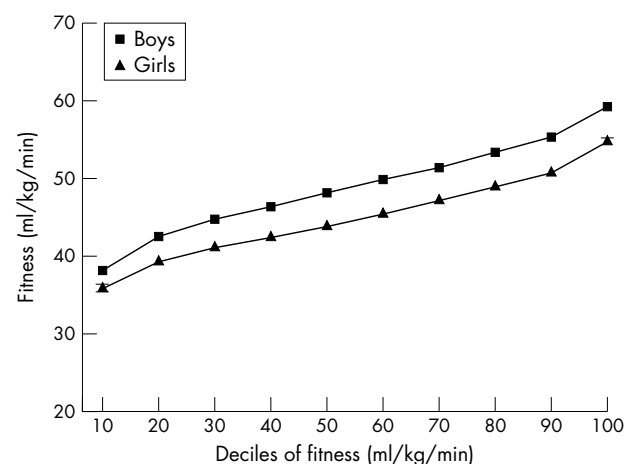


Figure 1 Subjects were ranked into 10 centiles of VO₂MAX by sex. The mean VO₂MAX was calculated for boys and girls separately in each centile. Standard errors are so small that they are not visible.

Only a few studies in the literature report measured values of VO₂MAX for children less than 8 years of age and only for a small number of children.^{8, 25–27} Mean values of 39–53 ml/kg/min have been reported. Differences in VO₂MAX between studies seem to be due, at least partly, to differences in protocol. The lowest mean value reported (39 ml/kg/min) came from a field test.²⁵ The highest mean values found were 53 and 52 ml/kg/min for 5–7 year old boys and girls respectively.²⁶ The highest values were calculated from the original data of Åstrand.²⁶ Testing was carried out on a treadmill with a continuous running protocol in a laboratory setting, and subjects were highly selected. Compared with earlier studies,^{8, 26} in which a difference in VO₂MAX relative to body weight of less than 2% between the sexes in prepubertal children was reported, our data show a considerable difference (8%). Rowland²³ suggested that at least some of the sex difference in children comes from the difference in fat percentage. In our study, the difference was 23%. As the boys were 3% heavier than the girls, this indicates that at least part of the sex difference in VO₂MAX is related to difference in body composition. The much smaller 2% difference we found in VO₂MAX relative to LBM between boys and girls supports this. Further, when allometric scaling was performed with LBM as scaling factor, the difference was only 1% and statistically insignificant. This implies that the sex difference in VO₂MAX is due mainly to a difference in body composition.

Vinet *et al*²⁸ and Rowland *et al*²⁹ came to different conclusions. Vinet *et al* found no difference in stroke volume and VO₂MAX after allometric scaling and stated that body composition alone (and not cardiac functional capacity) could account for the sex difference. Rowland *et al* found a difference between sexes even after allometric scaling and

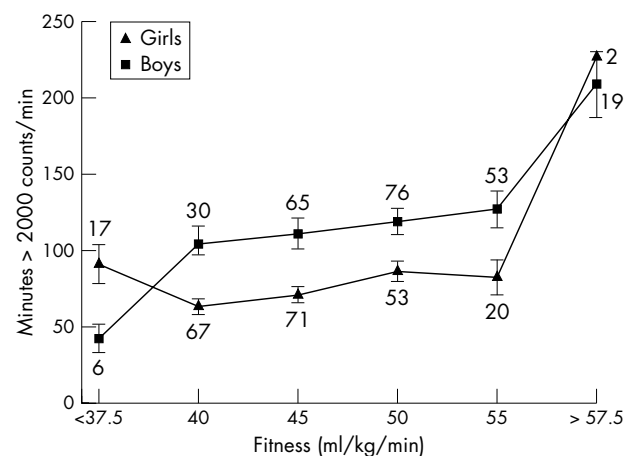


Figure 2 All subjects together were ranked into six groups of VO₂MAX relative to body weight. Physical activity level was plotted with SE for boys and girls with the same fitness level. Numbers of girls and boys are shown for each group.

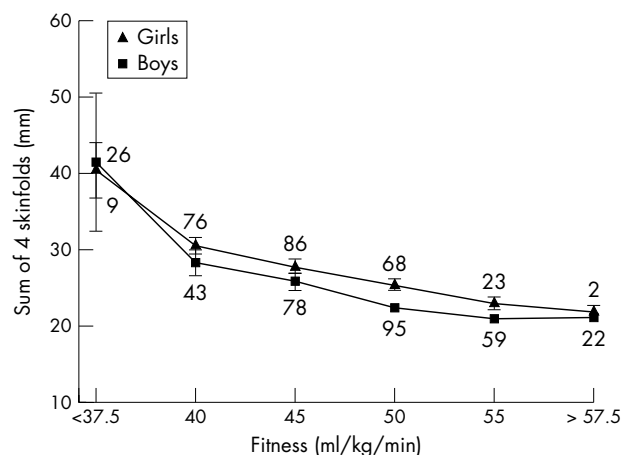


Figure 3 All subjects together were ranked into six groups of $\dot{V}O_{2\text{MAX}}$ relative to body weight. Sum of four skinfolds was plotted with SE for boys and girls with the same fitness level. Numbers of girls and boys are shown for each group.

stated that both body composition and cardiac functional capacity accounted for the difference between boys and girls. Further, Rowland *et al* found that anthropometric and aerobic physiological factors cannot entirely explain the magnitude of the sex differences.

Physical activity

We found a large sex difference (60%) in PA with regard to sustained activity, and sustained periods of PA explained 9% of the variance in $\dot{V}O_{2\text{MAX}}$ relative to body weight. It was tempting to claim that habitual PA may be part of the explanation for the difference in $\dot{V}O_{2\text{MAX}}$ between boys and girls. This is also the case, but even within the same stratum of $\dot{V}O_{2\text{MAX}}$ the boys had a higher PA level, so differences in PA seem only partly to explain the differences in $\dot{V}O_{2\text{MAX}}$. Sundberg³⁰ compared $\dot{V}O_{2\text{MAX}}$ (ml/kg/min) in sighted children and blind children, who are not able to participate in vigorous play, and found 22% and 26% higher values in sighted boys and girls respectively compared with blind children. From Sundberg's data, it could be expected that much of the difference between fitness levels in prepubertal children can be explained by differences in PA. It is therefore surprising in the present study that boys were more active than girls even within the same stratum of fitness. It is perhaps less surprising that boys have smaller SFS compared with girls in the same stratum of fitness. These two observations support the view of Rowland *et al*²⁹ that the higher level of fitness in boys is due to both the higher level of PA and the lower level of fatness.

Representativity

The number of children in this study is relatively high and represents 69% of all children in preschool in two suburban communities of Copenhagen. There were no significant differences between participants and non-participants in age, height, weight, and BMI in the study, so it seems safe to conclude that the participant cohort is representative of the two communities. Further, we have compared our sample with a sample from nationwide Danish studies with measurements of height and weight in children.^{31–32} The differences between these two studies and our study with regard to mean height and BMI are less than 1%. It seems reasonable to conclude that with regard to height and BMI our cohort is representative of Danish children aged 6–7 years. The strengths of this study are that a large number of

What is already known on this topic

No study has assessed population data in 6–7 year old children on aerobic fitness and physical activity using direct measurement of $\dot{V}O_{2\text{MAX}}$ and objective measurement of physical activity.

What this study adds

- Even in 6–7 year old children, boys had 8% higher fitness levels than girls, and the difference in fitness may mainly be explained by differences in physical activity level and body composition
- Moderate intensity physical activity may not influence physical fitness but still change body composition

subjects were examined using objective and direct measurements and that the sample was representative.

The number of rejected tests in this study is relatively high. Even though there were no differences in age and size between subjects with valid tests and rejected tests, the tendency was that younger and smaller children chose not to run or stopped prematurely. We also chose to reject the data for 57 children who actually ran a subjectively approved test, but did not have a $\dot{V}O_{2\text{MAX}}$ recording. In the end, 84% of the participating children completed a valid test, and we consider this satisfactory, especially as some authors have claimed that the reliability or validity of testing $\dot{V}O_{2\text{MAX}}$ in children aged less than 8 years is questionable.^{33–34} Our own validation of the protocol showed good reliability, and it seems reasonable to conduct reliable physiological testing on 6–7 year old children with satisfying results. It should be noted that care must be taken to carefully examine every child's measurements to ensure that the appropriate criteria have been met, and to discard non-valid tests. In conclusion, the absolute levels of $\dot{V}O_{2\text{MAX}}$ relative to body weight were 48.5 ml/kg/min and 44.8 ml/kg/min in girls, $\dot{V}O_{2\text{MAX}}$ is larger in boys than girls (+11%), also when related to body mass (+8%) and to LBM (+2%). Most of the difference in $\dot{V}O_{2\text{MAX}}$ relative to body mass was due to the greater percentage body fat in girls and lower level of PA. When boys and girls with the same $\dot{V}O_{2\text{MAX}}$ were compared, boys engaged in more minutes of at least moderately intense activity and had a smaller SFS.

Authors' affiliations

S Eiberg, H Hasselström, V Grønfeldt, K Froberg, J Svensson, Institute for Exercise and Sport sciences, University of Copenhagen, Denmark
L B Andersen, Norwegian University of Sport and Physical Education, Oslo, Norway

Competing interests: none declared

REFERENCES

- 1 Wedderkopp N, Froberg K, Hansen HS, *et al*. Secular trends in physical fitness and fatness in Danish 9-year old girls and boys. Odense School child Study and Danish substudy of The European Youth Heart Study. *Scand J Med Sci Sports Exerc* 2004;**14**:1–6.
- 2 Berenson GS, Wattigney WA, Tracy RE, *et al*. Atherosclerosis of the aorta and coronary arteries and cardiovascular risk factors in persons 6 to 30 years and studied at necropsy (the Bogalusa Heart Study). *Am J Cardiol* 1992;**70**:851–8.
- 3 Hasselström H, Hansen SE, Froberg K, *et al*. Physical fitness and physical activity during adolescence as predictors of cardiovascular disease risk in young adulthood. *Int J Sports Med* 2002;**23**:S27–31.

- 4 **Webber LS**, Voors AW, Srinivasan SR, *et al.* Occurrence in children of multiple risk factors for coronary artery disease: the Bogalusa Heart Study. *Prev Med* 1979;**8**:407–18.
- 5 **Wedderkopp N**, Froberg K, Hansen HS, *et al.* Cardiovascular risk factors cluster in children and adolescents with low physical fitness. *Pediatr Exerc Sci* 2003;**15**:419–22.
- 6 **Armstrong N**, Williams J, Balding J, *et al.* The peak oxygen uptake of British children with reference to age, sex and sexual maturity. *Eur J Appl Physiol* 1991;**62**:369–75.
- 7 **Cooper DM**, Weiler-Ravell D, Whipp BJ, *et al.* Growth-related changes in oxygen uptake and heart rate during progressive exercise in children. *Pediatr Res* 1984;**18**:845–51.
- 8 **Lemura LM**, von Duvillard SP, Cohen SL, *et al.* Treadmill and cycle ergometry testing in 5- to 6-year-old children. *Eur J Appl Physiol* 2001;**85**:472–8.
- 9 **McMiken DF**. Maximum aerobic power and physical dimensions of children. *Ann Hum Biol* 1976;**3**:141–7.
- 10 **Wong TW**, Yu TS, Wang XR, *et al.* Predicted maximal oxygen uptake in normal Hong Kong Chinese schoolchildren and those with respiratory diseases. *Pediatr Pulmonol* 2001;**31**:126–32.
- 11 **Shuleva KM**, Hunter GR, Hester DJ, *et al.* Exercise oxygen uptake in 3- through 6-year-old children. *Pediatr Exerc Sci* 1990;**2**:130–9.
- 12 **Andersen LB**, Henckel P, Saltin B. Maximal oxygen uptake in Danish adolescents 16–19 years of age. *Eur J Appl Physiol* 1987;**56**:74–82.
- 13 **Knutgen HG**. Aerobic capacity of adolescents. *J Appl Physiol* 1968;**22**:655–8.
- 14 **Zhu S**, St Onge MP, Heshka S, *et al.* Lifestyle behaviors associated with lower risk of having the metabolic syndrome. *Metabolism* 2004;**53**:1503–11.
- 15 **Brage S**, Wedderkopp N, Franks PW, *et al.* Reexamination of validity and reliability of the CSA monitor in walking and running. *Med Sci Sports Exerc* 2003;**35**:1447–54.
- 16 **Krahenbuhl GS**, Skinner JS, Kohrt WM. Developmental aspects of maximal aerobic power in children. *Exerc Sport Sci Rev* 1985;**13**:503–38.
- 17 **Armstrong N**, Kirby BJ, McManus AM, *et al.* Aerobic fitness of prepubescent children. *Ann Hum Biol* 1995;**22**:427–41.
- 18 **Armstrong N**, Welsman J. *Young people and physical activity*. London: Health Education Authority, 1997.
- 19 **De Lorenzo A**, Bertini I, Candeloro N, *et al.* Comparison of different techniques to measure body composition in moderately active adolescents. *Br J Sports Med* 1998;**32**:215–19.
- 20 **Dezenberg CV**, Nagy TR, Gower BA, *et al.* Predicting body composition from anthropometry in pre-adolescent children. *International Journal of Obesity* 1999;**23**:253–9.
- 21 **Ekelund U**, Sjöström M, Yngve A, *et al.* Physical activity assessed by activity monitor and doubly labeled water in children. *Med Sci Sports Exerc* 2001;**33**:275–81.
- 22 **Jensen K**, Jorgensen S, Johansen L. A metabolic cart for measurement of oxygen uptake during human exercise using inspiratory flow rate. *Eur J Appl Physiol* 2002;**87**:202–6.
- 23 **Rowland TW**. *Developmental exercise physiology*. Champaign, IL: Human Kinetics, 1996:1–269.
- 24 **Altman DG**. *Practical statistics for medical research*. London: Chapman & Hall/CRC, 1991:36–7.
- 25 **Yoshida T**, Ishiko I, Muraoka I. Effects of endurance training on cardiorespiratory functions of 5-year old children. *Int J Sports Med* 1980;**1**:91–4.
- 26 **Åstrand P-O**. Experimental studies of physical working capacity in relation to age and sex. 1952:1–171.
- 27 **Robinson S**. Experimental studies of physical fitness in relation to age. *Arbeitsphysiologie* 1938;**10**:251–323.
- 28 **Vinet A**, Mandigout S, Nottin S, *et al.* Influence of body composition, hemoglobin concentration, and cardiac size and function of gender differences in maximal oxygen uptake in prepubertal children. *Chest* 2003;**124**:1494–9.
- 29 **Rowland T**, Goff D, Martel L, *et al.* Influence of cardiac functional capacity on gender differences in maximal oxygen uptake in children. *Chest* 2000;**117**:629–35.
- 30 **Sundberg S**. Maximal oxygen uptake in relation to age in blind and normal boys and girls. *Acta Paediatr Scand* 1982;**71**:603–8.
- 31 **Rasmussen S**, Petersen TA, Madsen M. [Body height of 6–15-year-old school children measured in the period 1986/1987 to 1996/1997. Compared with Danish measurements in 1971/1972]. *Ugeskr Laeger* 2002;**164**:5011–15.
- 32 **Petersen A-G**, Rasmussen S, Madsen M. Danske skolebørns BMI målt i Perioden 1986/87–1996/97 sammenlignet med danske målinger fra 1971/72. *Ugeskr Laeger* 2002;**164**:5006–10.
- 33 **Malina RM**, Bouchard C. *Growth, maturation and physical activity*. Champaign, IL: Human Kinetics, 1991.
- 34 **Armstrong N**, Welsman JR. Assessment and interpretation of aerobic fitness in children and adolescents. *Exerc Sport Sci Rev* 1994;**22**:435–76.

11th European Forum on Quality Improvement in Health Care

26–28 April 2006, Prague, Czech Republic
 For further information please go to: www.quality.bmjpg.com
 Book early to benefit from a discounted delegate rate

RESEARCH ARTICLE

Physical Fitness in Spanish Schoolchildren Aged 6-12 Years: Reference Values of the Battery EUROFIT and Associated Cardiovascular Risk

ROBERTO GULÍAS-GONZÁLEZ, PhD^a MAIRENA SÁNCHEZ-LÓPEZ, PhD^b ÁNGEL OLIVAS-BRAVO, BSc Ed^c MONTSERRAT SOLERA-MARTÍNEZ, PhD^d
VICENTE MARTÍNEZ-VIZCAINO, MD, PhD^e

ABSTRACT

BACKGROUND: Physical fitness is considered an important indicator of health in children. The aims of this study were to (1) provide sex- and age-specific EUROFIT battery levels of fitness in Spanish children; (2) compare Spanish children's fitness levels with those of children from other countries; and (3) determine the percentage of Spanish children with cardiovascular risk associated with low cardiorespiratory fitness (CRF).

METHODS: Physical fitness was assessed using the EUROFIT tests in 1725 children, aged 6 to 12, from Castilla-La Mancha, Spain. We derived specific values for physical fitness using LMS method. FITNESSGRAM 2010 criteria were used to estimate the percentage of children with cardiovascular risk associated with low CRF.

RESULTS: Boys scored higher in all the physical fitness tests, except for the flexibility test. Physical fitness improved as age increased, except for flexibility, which worsened in boys, and VO_{2max} , which decreased in both sexes. The prevalence of boys and girls with cardiovascular risk associated to low CRF was 13% and 26%, respectively.

CONCLUSIONS: Specific fitness test scores for children and adolescents can represent the fitness status of schoolchildren accurately. Schools need to make efforts to improve the fitness level of the schoolchildren to prevent cardiovascular risk.

Keywords: child and adolescent health; physical fitness and sport; public health.

Citation: Gulías-González R, Sánchez-López M, Olivas-Bravo Á, Solera-Martínez M, Martínez-Vizcaino V. Physical fitness in Spanish schoolchildren aged 6-12 years: reference values of the battery EUROFIT and associated cardiovascular risk. *J Sch Health*. 2014; 84: 625-635.

Received on October 25, 2012

Accepted on January 5, 2014

Physical fitness is considered an important indicator of health in children and adolescents.^{1,2} High levels of cardiorespiratory fitness (CRF) in childhood are associated with increased cardiovascular, skeletal, and mental health,^{1,2} and improvements in muscular strength levels from childhood to adolescence are

associated with a decrease in total adiposity.² Although less studied, speed/agility is strongly related to bone mineral density and bone mass accumulation in later stages of life.²⁻⁴ Low levels of physical fitness in childhood are associated with an increased risk of cardiovascular disease in adulthood.³⁻⁵ Furthermore,

^aAssistant Professor, (Roberto.Gulias@uclm.es), Faculty of Education, University of Castilla-La Mancha, Ronda de Calatrava, 3, 13071 Ciudad Real, Spain.

^bAssistant Professor, (Mairena.Sanchez@uclm.es), Faculty of Education, University of Castilla-La Mancha, Ronda de Calatrava, 3, 13071 Ciudad Real, Spain.

^cAssistant Professor, (Angel.Olivas@uclm.es), Faculty of Education, University of Castilla-La Mancha, Ronda de Calatrava, 3, 13071 Ciudad Real, Spain.

^dAssistant Professor, (Montserrat.Solera@uclm.es), Social and Health Care Research Center; University of Castilla-La Mancha, Santa Teresa Jornet s/n, 16071 Cuenca, Spain.

^eProfessor, (Vicente.Martinez@uclm.es), Social and Health Care Research Center, University of Castilla-La Mancha, Santa Teresa Jornet s/n, 16071 Cuenca, Spain.

Address correspondence to: Roberto Gulías-González, Assistant Professor, (Roberto.Gulias@uclm.es), Faculty of Education, University of Castilla-La Mancha, Ronda de Calatrava, 3, 13071 Ciudad Real, Spain

The authors of this study express their appreciation and gratitude to all the schools, students, and teachers who participated in this study voluntarily. Also to all members involved in the fieldwork for their effort and enthusiasm. This study was funded by the Ministry of Education and Science through aid for the development of cooperation projects in the field of innovation and research among the faculty and nonuniversity teachers of the autonomous community of Castilla-La Mancha, Spain (Resolution 08/07/2009, the Deputy Ministry of Education of the Regional Government of Castilla-La Mancha). Data about physical fitness smoothed centile values for the EUROFIT battery of tests by sex and age in the schoolchildren are available from the corresponding author.

the level of physical fitness in childhood is a predictor of physical fitness in adult life;^{2,5-8} therefore, to include measurements of physical fitness related to health in the education and health systems is clearly justified.

In Spain, the main drivers behind the alarming trend of obesity prevalence in schoolchildren are richly debated. A possible way to try to clarify this issue is to compare the fitness levels of Spanish children with those of children in other countries where the prevalence of obesity is appreciably lower. Considering the consistently described relationship between daily physical activity and fitness levels, if the levels of fitness of Spanish children are lower than the levels in countries with lower obesity prevalence, such as northern European countries, the sedentary lifestyle could be considered as the main driver of the current obesity epidemic.

To assess physical fitness, updated values for the target population are necessary. Owing to methodological differences in the protocols of assessment of physical fitness, comparing and interpreting the results between and among different population studies is complicated.^{9,10} EUROFIT battery has proven to be a simple, standardized, and practical instrument, with good reliability and validity.¹¹⁻¹⁴

Traditionally, the program FITNESSGRAM has provided valid and reliable cutoff points for identifying children with future cardiovascular risk according to their level of CRF.^{15,16} However, improved standards developed with advanced statistical techniques (Receiver Operating Characteristic-ROC-curves) have recently been published. These new criteria allow the categorization of children into 3 fitness levels (Healthy Fitness Zone, Higher Risk, and Some Risk), thereby providing more specific and clearer information than the previous 2 categories (Healthy Fitness Zone and Needs Improvement Zone) established by traditional standards.^{17,18} Placing individuals and groups in percentiles and categories allow to identify individuals with low fitness levels and to implement interventions to promote healthy behaviors in order to prevent future cardiovascular risk. This study aimed to (1) assess the physical fitness of children, aged 6-12, from Castilla-La Mancha, Spain, according to sex and age; (2) compare our results with those of other European and international studies conducted in the same population group; and (3) estimate the prevalence of students who present future cardiovascular risk associated with low CRF, in schoolchildren aged 10-12 in Castilla-La Mancha, Spain.

METHODS

Participants

This is a multicenter, cross-sectional observational study including 1725 both sexes, aged 6-12 years, primary schoolchildren from 75 public and private

schools randomly selected from the 778 schools in Castilla-La Mancha by multistage sampling. In each school, 2 boys and 2 girls were randomly selected per class, 1st to 6th grades. A sample size of 629 subjects was estimated by using EpiData 3.1 software (EpiData Association, Odense, Denmark), considering a prevalence of cardiovascular risk associated with low CRF of 18%,⁹ a confidence level of 95%, and 3% accuracy. Because we intended to maintain that accuracy in both sexes, and considering a nonresponse rate of 30%, the minimum sample size was estimated in 1635 students.

Data from 2 population-based samples (Latvia and Australia) were used to compare data from Spanish students with data belonging to other countries. These samples were selected because of both of them employed the same fitness tests used in our study, and were also large samples from the same age range.

Instruments and Procedure

Measurements were carried out on the premises of each school between April and May 2010. Tests were carried out by researchers with Bachelor of Science degrees in physical activity and education, and by physical education teachers who had previously completed a 2-day training course to standardize measurements.

Evaluation of physical fitness. In the gymnasium of each school, physical fitness was assessed using the 9 tests included in the EUROFIT battery:^{11,12} Flamingo Balance Test, Plate Tapping, Sit and Reach Test, Standing Broad Jump Test, Handgrip Test, Sit-ups in 30 seconds, Bent Arm Hang, 10 × 5 m Shuttle Run, 20-m Shuttle Run Test.

First, each school was contacted to obtain the approval of the director and the school board. None of the schools invited refused to participate. Subsequently, newsletters explaining the objectives of the study, and requesting written consent for students' participation, were sent to the parents of the selected students. The design of the study was conducted under the principles established in the Declaration of Helsinki, and information management was conducted according to the provisions laid forth in Spanish law (Law 15/1999 of 13 December on the Protection of Personal Data).

Data Analysis

The adjustment to normal distribution of the different variables was evaluated both by graphs and by the Kolmogorov-Smirnov test. Fitness levels by age (years) and sex were presented as mean and standard deviation. Sex differences on fitness levels were tested by using Student-*t* statistic.

To provide smoothed centile curves (P_{10} , P_{25} , P_{50} , P_{75} , and P_{90}) for schoolchildren, physical fitness data were analyzed by maximum penalized likelihood

using the LMS statistical method by sex and age^{19,20} (LmsChartMaker software V. 2.3; by Tim Cole and Huiqi Pan). The LMS method does not work with 0 values. Because a number of schoolchildren scored 0 in the bent arm hang test, we estimated centile values for this test using standard procedures instead of the LMS method.

The cardiovascular risk associated with low CRF for children was calculated using the new cutoff points proposed in 2010 by the Cooper Institute's FITNESSGRAM group.^{17,18} Only subjects in the range of 10-12 years old were taken into account for this calculation, since there are no standards for subjects younger than 10 years of age. We used the statistical chi-square to determine the existence of differences by sex in the prevalence of cardiovascular risk associated with low CRF.

IBM SPSS 19 Statistics software was used, except in the LMS method calculations, and the significance level was set at $p \leq .05$.

RESULTS

Table 1 shows the results of the physical fitness tests by age and sex. Boys scored higher than girls in all the fitness tests, except in the flexibility test. Overall, the mean scores for the physical fitness variables increase with age for boys and girls, except for flexibility, which decreases in boys and remains stable in girls, and the levels of VO_{2max} (mL/kg/minutes) which decrease in both sexes. The smoothed centile curves (P_{10} , P_{25} , P_{50} , P_{75} , and P_{90}) for different fitness tests for both sexes, by age, were in the same direction as in Table 1 (Figures 1-3).

Figure 4 depicts the smoothed centile curves (P_{10} , P_{50} , and P_{90}) of the main components of physical fitness in Spanish, Latvian, and Australian children, by sex and age. In general, Latvian and Australian schoolchildren scored higher than Spanish children in P_{90} . In P_{50} and P_{10} , fitness tests figures were similar to those of Latvian children; however, Spanish children scored lower in strength and flexibility, in both sexes, compared with their Australian counterparts.

Cardiovascular risk associated with their current CRF was 26% for girls (5% higher risk, 21% some risk) and 13% for boys (2.8% higher risk, 10.2% some risk).

DISCUSSION

This study provides, through smoothed centiles, updated values for different tests of physical fitness in schoolchildren aged 6 to 12, from Castilla-La Mancha, Spain. In our study, boys show better values of physical fitness than girls in all components of fitness, except for flexibility. As age increases, students improve in all the fitness levels, except for flexibility

and VO_{2max} (mL/kg/minutes). Older boys show less flexibility levels than younger; in girls, after puberty the flexibility levels are stable. In both, boys and girls, VO_{2max} (mL/kg/minutes) values declined with age (Figures 1-3).

As in other national,²¹ European,^{13,22-24} and international studies,^{25,26} boys showed higher levels of fitness than girls, except for flexibility. It has been reported that, before puberty, differences in fitness between sexes are absent or negligible;²⁷⁻³⁰ however, other studies have found that these sex differences could be due to morphological characteristics (different somatotype)³¹ and physiological traits (eg, differences in mechanical efficiency and/or fractional utilization).³²

In our study, as in others,^{22,25,26} musculoskeletal, respiratory, and motor abilities increased with age during the school-age years (Figures 1 and 3). It seems paradoxical, however, that in the case of VO_{2max} (mL/kg/minutes), calculated from the data gathered from the 20-m shuttle run test, the younger schoolchildren of our study showed the highest levels (Table 1 and Figure 2). In our opinion, the reason for these unexpected results is that although the absolute VO_{2max} (mL) increases with age,^{27,33} the increase in adiposity is responsible for the decrease in VO_{2max} , expressed in relation to body weight. This has been observed in Spanish subjects of this age group,³⁴ as well as described in other studies.^{32,35} The decline or stability of flexibility found in the children of our study (Figure 2) has been described in other studies,^{13,22,25,26} and is not surprising because flexibility, if not specifically maintained, diminishes with age.³⁶

Upon comparing our data with those taken from Australian and Latvian schoolchildren, it was striking to see that Spanish students show lower fitness levels in the highest categories (P_{90}); in other words, when looking at those children who performed highest, Spanish schoolchildren score worse on all the tests, by age and sex, than their Australian and Latvian counterparts. For all other centiles (P_{50} and P_{10}), our data are similar to those observed in Latvian children.¹³ Compared to data taken from Australian children, Spanish children show lower muscle strength and flexibility, with the differences between the countries more pronounced in boys.²⁵

Because it is unlikely that substantial genetic differences exist between our sample and Australian and Latvian children, it is probable that differences in physical activity explain the lower physical fitness values that we observed. In one study describing physical activity patterns in Spanish children of 9 years of age, boys were shown to perform an average of 81.8 minutes/day of moderate-vigorous physical activity (MVPA), and girls an average of 57.51 minutes/day.³⁷ Meanwhile, in Australia,

Table 1. Physical Fitness Scores (Mean $\pm$ SD) of EUROFIT Battery for Spanish Schoolchildren by Age and Sex

Age (Years)	Muscular Strength					Cardiorespiratory Endurance (20 m Endurance Shuttle Run Test)				Speed		Total Body Balance
	Muscular Endurance		Maximal Strength	Explosive Strength		Flexibility	Cardiorespiratory Endurance (20 m Endurance Shuttle Run Test)			Shuttle Run Test (10 x 5) (seconds) [‡]	Plate Tapping Test (seconds) [‡]	
	Bent Arm Hang Test (s)	Sit-Ups Test (Repeats)	Handgrip Test (kg)	Broad Jump Test (cm)	Sit and Reach Test (cm)		VO _{2max} (mL/kg/ minutes) [†]	(stage)				
Boys (N = 868)												
6	5.6 ± 6.5*	12.3 ± 6.1	10.2 ± 2.5	106.5 ± 16.3**	16.3 ± 5.1		2.1 ± 1.3**	49.3 ± 2.8**	25.2 ± 3**	22.9 ± 3.9	20.3 ± 6.6	
7	6.6 ± 7*	14.9 ± 6.1*	12.1 ± 2.8**	113.6 ± 19.1**	16.2 ± 5.2		2.5 ± 1.4*	48.2 ± 3.1*	24 ± 3.5*	20.5 ± 3.6	17.1 ± 7.6*	
8	8 ± 8.7*	16.7 ± 5.9*	13.9 ± 2.5**	122.4 ± 17.9**	15.2 ± 5.3		3 ± 1.9*	47.4 ± 4.3*	23 ± 2.5*	18.3 ± 3.1	14.8 ± 7*	
9	8.4 ± 8.6*	18 ± 5.5*	15.7 ± 3.6**	125.1 ± 20.5**	15.3 ± 6.5		3.5 ± 2**	46.8 ± 4.6**	22.5 ± 2.3**	17.1 ± 2.8	13.1 ± 7.4	
10	9.6 ± 10.8**	19.7 ± 5**	17.1 ± 3.9*	131 ± 18.3**	12.8 ± 6.5		3.8 ± 1.9**	45.9 ± 4.5**	21.8 ± 2.3**	15.7 ± 2.5	11.4 ± 7.7	
11	10.4 ± 10**	20.5 ± 5.7*	18.8 ± 4	141.4 ± 23.5**	13.3 ± 6.1		4.5 ± 2.1**	45.9 ± 5.3**	21.4 ± 2.3**	14.7 ± 2.3	9.3 ± 6.7	
12	11.8 ± 11.6*	21.9 ± 6.5*	22.8 ± 5.6	148.9 ± 25.6**	12.5 ± 6.9		5.1 ± 2.1**	45.7 ± 5.4**	21 ± 3.2**	14 ± 2.3*	9.5 ± 6.3	
Girls (N = 857)												
6	3.8 ± 4.2	11.6 ± 6.2	9.5 ± 2.2	95.3 ± 19.2	18.6 ± 6*		1.5 ± 0.7	47.9 ± 1.5	26.2 ± 3.1	23.1 ± 4.1	19.3 ± 6.3	
7	4.5 ± 4.7	12.7 ± 5.9	10.8 ± 2.5	100.4 ± 14.9	17.6 ± 5.6*		2 ± 1	47.1 ± 2.1	25 ± 2.3	21 ± 3.6	13.8 ± 7.1	
8	5.6 ± 6.6	15 ± 5.4	12.5 ± 2.8	111.7 ± 16.6	18 ± 5.9**		2.3 ± 1.3	46 ± 2.8	24.1 ± 2.8	18.3 ± 2.9	12.2 ± 7.6	
9	5.5 ± 5.2	16 ± 5.4	14.1 ± 2.9	113.2 ± 17.9	16.7 ± 6.7		2.3 ± 1.3	44.2 ± 3	23.9 ± 2.4	17.3 ± 2.9	11.6 ± 6.8	
10	4.5 ± 5.9	16.4 ± 6	16 ± 3.6	118.8 ± 19.5	17.8 ± 7**		2.8 ± 1.4	43.3 ± 3.4	23.3 ± 2.8	15.6 ± 2.4	10.5 ± 7.8	
11	6.4 ± 6.9	18.4 ± 5.3	19 ± 4.2	129 ± 21.2	17.3 ± 6.8**		3.1 ± 1.5	42.3 ± 3.8	22.5 ± 2.4	14.3 ± 1.9	8.7 ± 6.3	
12	7 ± 8.2	19.4 ± 4.4	21.5 ± 4.4	132 ± 24.4	19 ± 7.1**		3.8 ± 1.7	42.4 ± 4.3	22.3 ± 2.3	13.4 ± 1.9	9.4 ± 7	

*Mean differences by sex statistically significant ($p \leq .05$).**Mean differences by sex statistically significant ($p \leq .001$).[†]VO_{2max}, maximal oxygen consumption, calculated as: VO_{2max} = 31.025 + 3.238S - 3.248A + 0.1536SA, where A, age; S, final speed (8 + 0.5 $\times$ last stage completed).[‡]Lower values indicate better performance.

Figure 1. Smoothed Centile Curves (From Top to Bottom: P_{90} , P_{75} , P_{50} , P_{25} , and P_{10}) for (a) Bent Arm Hang (s), (b) Sit-Up Test (repetitions), (c) Handgrip Test (kg), and (d) Broad Jump Test (cm)

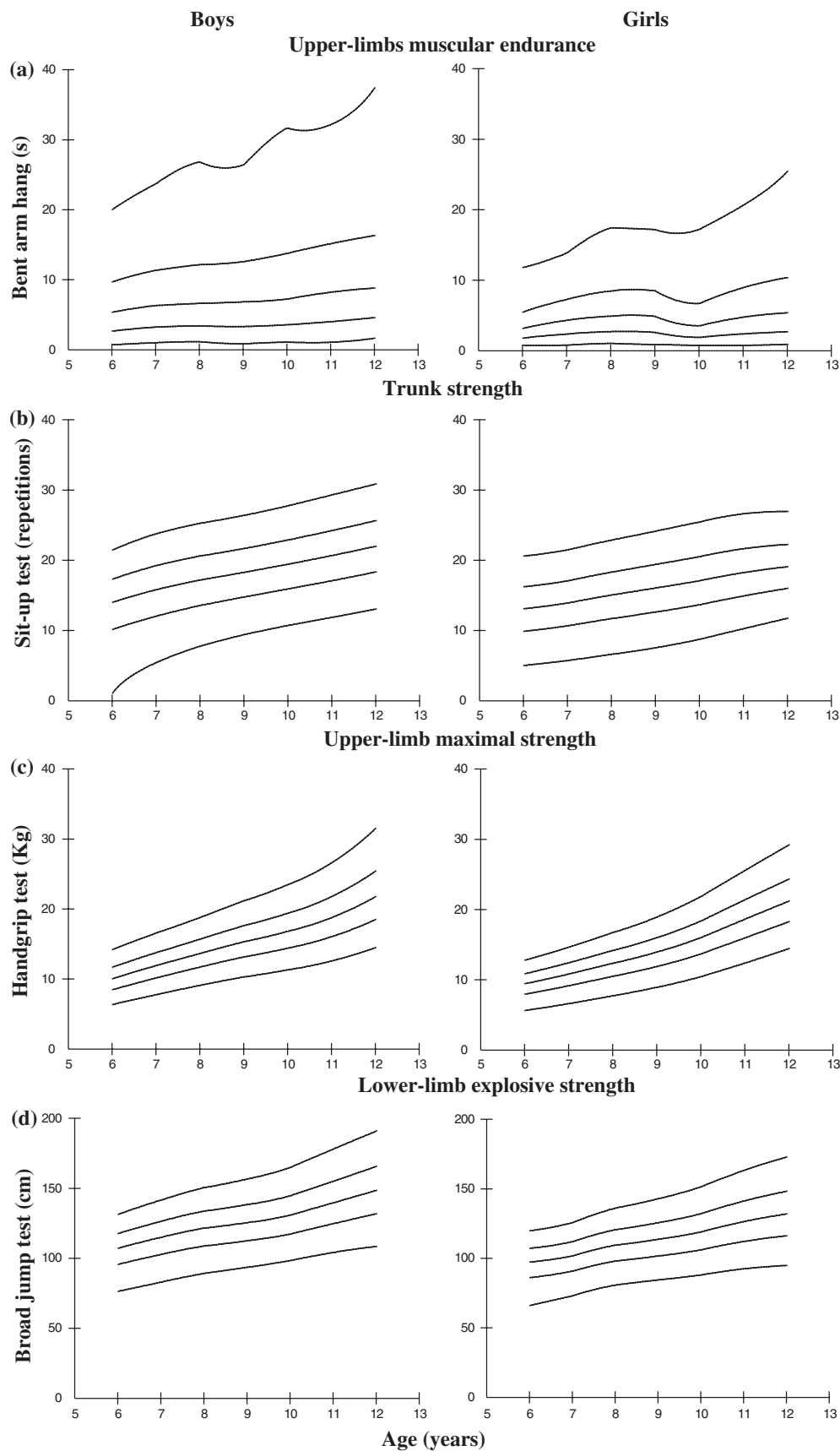
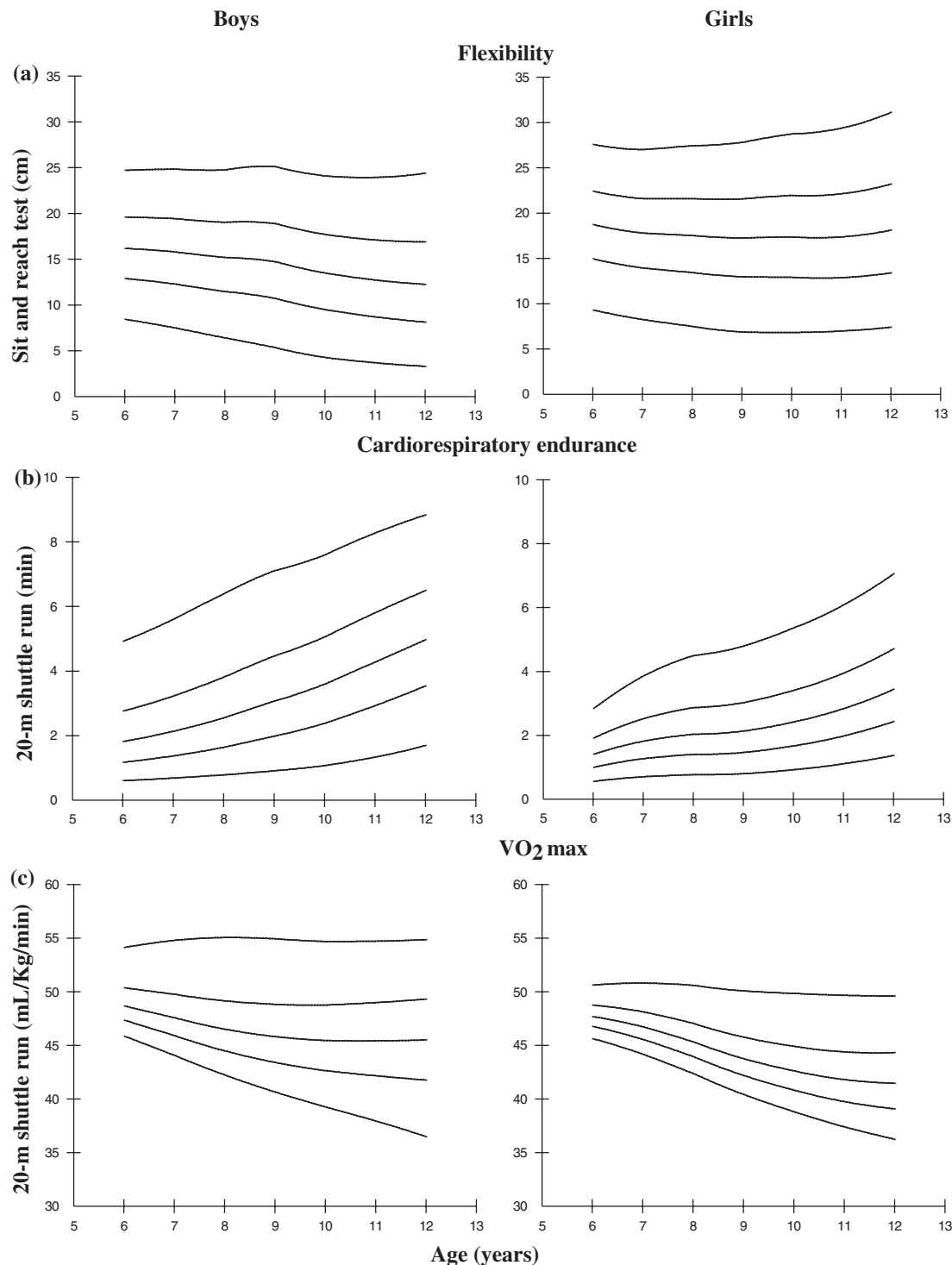


Figure 2. Smoothed Centile Curves (From Top to Bottom: P₉₀, P₇₅, P₅₀, P₂₅, and P₁₀) for (a) Sit and Reach Test (cm), (b) 20-m Shuttle Run (min), and (c) 20-m Shuttle Run (mL/kg/minutes)

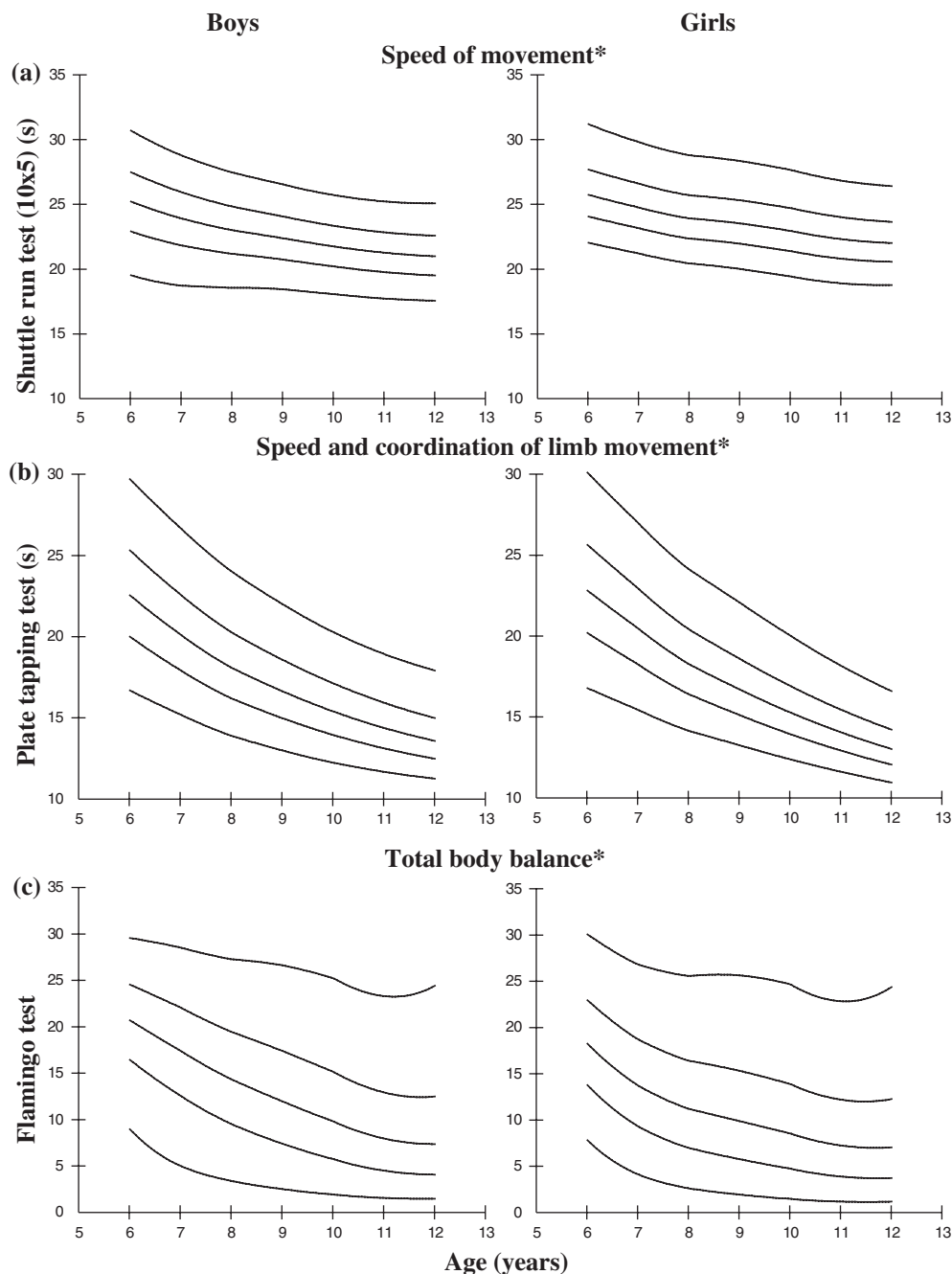


127.1 minutes/day is the average for boys and 100.9 minutes/day for girls.³⁸ Similarly, the HBSC (Health Behaviors in School-aged Children) study in children aged 11 to 15 described that 54% of Latvian boys and 40% of Latvian girls performed 1 hour/day

of MVPA, 5 days a week, compared to 46% of boys and 34% of girls in Spain.³⁹

Our study can inform the physical fitness levels in schoolchildren (ages 6 to 12) from the center of Spain, and estimates the corresponding smoothed centiles

Figure 3. Smoothed Centile Curves (From Top to Bottom: P₉₀, P₇₅, P₅₀, P₂₅, and P₁₀) for (a) Shuttle Run Test (10 × 5) (s), (b) Plate Tapping Test (s), and (c) Flamingo Test (Number of Falls). *Lower scores indicate better performance

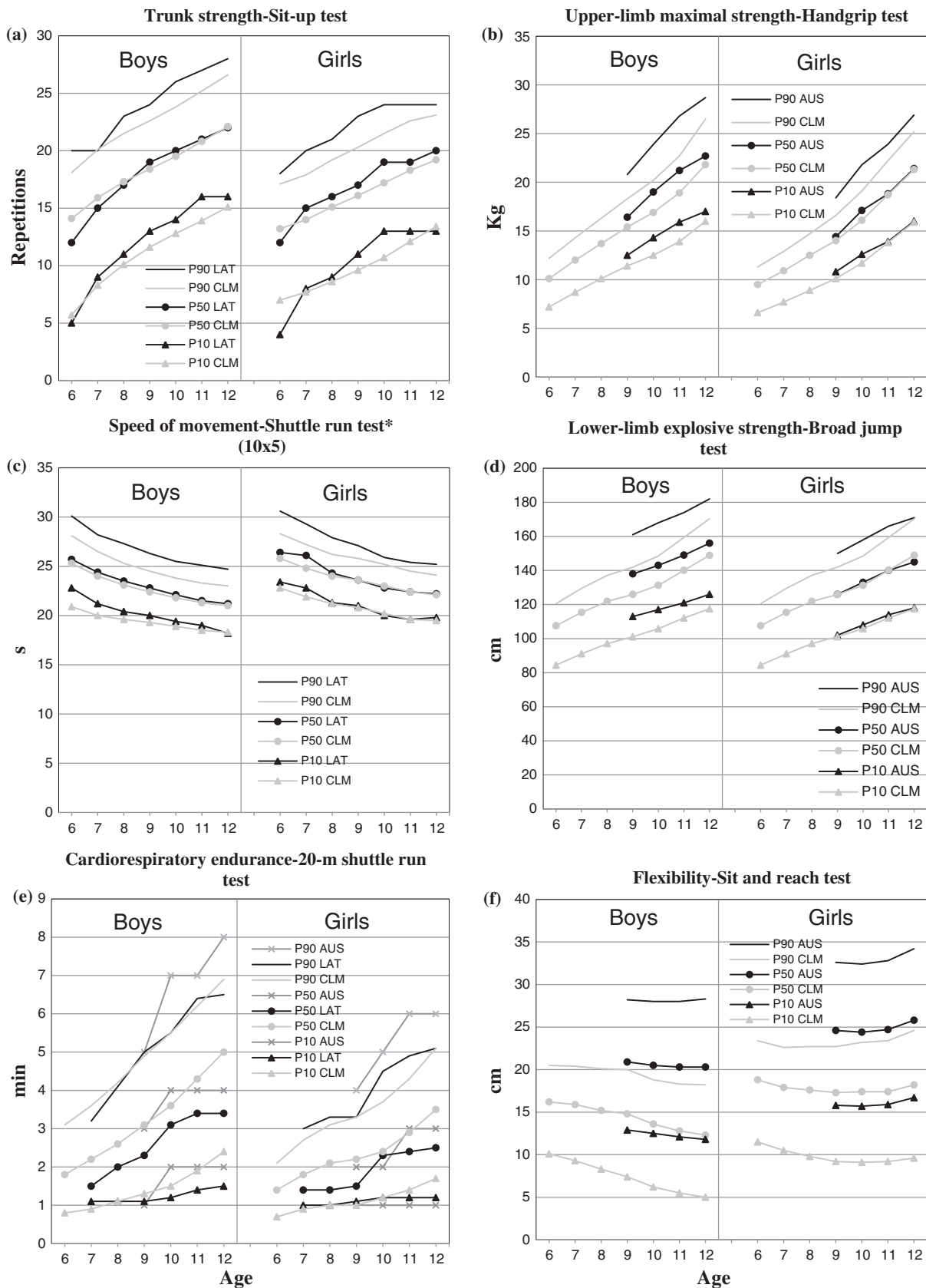


according to age and sex (Figures 1-3) that could be used as reference values to identify children with low fitness at school. According to other authors,¹⁰ falling below the lowest decile (P₁₀) should be considered as a warning signal.

In terms of population, having a fitness level below the 5th centile is usually considered an indicator of future cardiovascular risk.^{40,41} Other authors establish 5 risk categories based on the quintile in which one's

fitness level lies.⁹ From a clinical point of view, it is more useful to establish fitness categories that are associated with future cardiovascular risk, such as those proposed by the Cooper Institute through data obtained from the FITNESSGRAM program. With those data, 2 categories, or "zones" were established, according to aerobic capacity: a Needs Improvement Zone (NIZ)—further divided into "some risk" and "high risk"—and a Healthy Fitness Zone (HFZ).

Figure 4. Physical Fitness Centile Curves (P_{10} , P_{50} , and P_{90}) of 6 EUROFIT Tests From This Study (Castilla-La Mancha, Spain; CLM) and From 2 Countries (AUS, Australia and LAT, Latvia). *Lower scores indicate better performance



Considering jointly the NIZ areas, 26% of girls and 13% of boys ages 10-12 from our study are at levels of aerobic capacity associated with future cardiovascular risk. We observe that the percentage of girls that show future cardiovascular risk is twice that of boys, whereas maintaining this relationship in the 2 risk subzones. According to these data, 1 in 4 girls and 1 in 10 boys have a high likelihood of cardiovascular disease in adulthood if their CRF does not improve.

Our data are similar to those found in American children who, when the new FITNESSGRAM cutoff points were used, showed that 27.2% of girls and 19% of boys have cardiovascular risk associated with low CRF.¹⁷ Our data are also similar to those found in Portugal, where 28.3% of girls and 18.5% of boys showed cardiovascular risk associated with low CFR, according to the Cooper Institute criteria of 1999.⁴² Our data contrast with those of the EYHS (European Youth Heart Study), in which more than 40% of Swedish and Estonian schoolchildren aged 9 to 10 had levels of aerobic capacity associated with cardiovascular risk.⁴³ In view of the low prevalence of overweight, and the high levels of physical activity of Swedish and Estonian children,^{44,45} the estimates of low aerobic capacity reported in the EYHS are surprising. We believe that, in addition to the fact that different cutoff points were used in the EYHS, these differences in the estimates of the prevalence of NIZ could be attributed to differences in the methodology used for estimating VO_{2max} . In our study, Léger's equation using data from the 20-m shuttle run test was used for the indirect estimation of VO_{2max} , whereas in the EYHS was directly assessed by means of a maximal cycle ergometer fitness test. To our knowledge, there are no prevalence data of cardiovascular risk associated with low CRF in Spanish children aged 10-12 using the new criteria of Cooper Institute.⁴⁶ Casajús et al³⁰ working with children in Aragon, a region of northern Spain, reported that 9.1% of boys and 4.8% of girls do not reach the minimum levels recommended for healthy cardiovascular fitness. These percentages are clearly lower than those reported in our study. One possible explanation lies in the use of different cutoff points to categorize children at risk of future cardiovascular disease: 42 and 38 mL/kg/minutes for boys and girls respectively vs, in our study, 37.3 mL/kg/minutes for boys and girls aged 10-11, and 37.6 mL/kg/minutes for boys aged 12 and 37.0 mL/kg/minutes for girls aged 12. In the HELENA study, a European study including adolescents from 9 countries, 39% of boys and 43% of girls had high probability of future cardiovascular disease because of low CRF.¹⁰ These high figures during adolescence coincide with the gradual abandonment of physical/athletic activity that occurs among schoolchildren with the arrival of adolescence, which may explain the differences between children and adolescents.

Limitations

One of the major limitations of this study is that it includes children from only a single region of Spain; therefore, inferences to all Spanish children should be made cautiously. However, the sample size, the inclusion of both rural and urban schools, and the similarity between fitness values found in our study and those found in other studies conducted in different geographical areas of Spain,^{21,30,47,48} all suggest that our data may be representative of the set of Spanish children aged 6-12. However, because our study is a cross-sectional one, it may be that a cohort effect has occurred, and as a consequence, our estimations of fitness levels could not be extrapolated to previous cohorts. Finally, a decrease in the prevalence of overweight in Spanish schoolchildren would likely improve CRF in terms of VO_{2max} (mL/kg/minutes), and thus lower future cardiovascular risk.

Conclusions

Our study shows the physical health levels of Spanish schoolchildren aged 6-12 of Castilla-La Mancha, Spain, which could be useful in identifying children with low fitness and developing strategies to improve it. The physical fitness of Spanish children is worse than that of children from other countries. Within Spain, 1 in 4 girls and 1 in 10 boys has a level of CRF indicative of future cardiovascular risk. The high prevalence of cardiovascular risk associated with low aerobic capacity and low levels of strength compared with other countries in our study highlights the need to make deliberate, carefully planned efforts in the Spanish schools to improve the physical fitness of schoolchildren.

IMPLICATIONS FOR SCHOOL HEALTH

With 1 in 5 schoolchildren exhibiting aerobic capacity levels associated with future cardiovascular risk, Spanish schoolchildren are less physically fit than their counterparts outside Spain. It has been suggested that these alarming figures are associated with low levels of physical activity in the population.³⁷ Because of the relationships between fitness and cardiovascular risk,⁴⁹ and between physical exercise and fitness⁵⁰ that have been described, it is necessary to implement physical activity programs focused on the improvement of aerobic capacity and muscular strength. Such programs would obviously require collaboration among schools, health workers, politicians, and families.

Schools are places where children spend a large amount of time, and thus, function as a setting in which to implement interventions to fight against sedentariness in infancy and, as a consequence, against cardiovascular disease in the coming years. Several strategies for promoting physical activity on schools

have demonstrated some effectiveness in improving cardiometabolic health;⁵¹ thus, the challenge at this moment is to implement these interventions in an environment of economic crisis, even if they have a proven acceptable cost effectiveness ratio.⁵² In Spain, this challenge has a pivotal role as 2 circumstances converge: first, the time allocated to physical exercise is declining in the school schedules, and second, the prevalence of sedentary behavior during leisure time is increasing.

To promote changes in students' lifestyles, education and health authorities need to work with school directors to evaluate the implementation of physical activity programs of moderate to vigorous intensity, which could take place during a recess or before or after school hours. These programs should include recreational activities, sports, dancing, and outdoor activities. It would be desirable for these programs to include interventions that reach beyond the school to improve other aspects of a healthy lifestyle, such as eating habits or commuting behavior.⁵³

We have gathered values for the health-related physical fitness of children aged 6-12 in Castilla-La Mancha, Spain. These values can be used both to assess students' fitness levels and to detect the students whose fitness levels are below a healthy minimum.

Human Subjects Approval Statement

The study was approved by the Ministry of Education and Science of the Regional Government of Castilla-La Mancha, and the Faculty of Education at the University of Castilla-La Mancha, Spain. All parents or guardians of the subjects participating in the study signed informed consent forms to participate in the study.

REFERENCES

1. Ruiz JR, Castro-Pinero J, Artero EG, Ortega FB, Sjöström M, Suni J, Castillo MJ. Predictive validity of health-related fitness in youth: a systematic review. *Br J Sports Med*. 2009;43(12):909-923.
2. Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: a powerful marker of health. *Int J Obes (Lond)*. 2008;32(1):1-11.
3. Castillo-Garzon MJ, Ruiz JR, Ortega FB, Gutierrez-Sainz A. A Mediterranean diet is not enough for health: physical fitness is an important additional contributor to health for the adults of tomorrow. *World Rev Nutr Diet*. 2007;97:114-138.
4. Vicente-Rodriguez G, Ara I, Perez-Gomez J, Serrano-Sanchez JA, Dorado C, Calbet JA. High femoral bone mineral density accretion in prepubertal soccer players. *Med Sci Sports Exerc*. 2004;36(10):1789-1795.
5. Janz KF, Dawson JD, Mahoney LT. Increases in physical fitness during childhood improve cardiovascular health during adolescence: the Muscatine Study. *Int J Sports Med*. 2002;23(suppl 1):S15-S21.
6. Boreham C, Twisk J, Neville C, Savage M, Murray L, Gallagher A. Associations between physical fitness and activity patterns during adolescence and cardiovascular risk factors in young adulthood: the Northern Ireland Young Hearts Project. *Int J Sports Med*. 2002;23(suppl 1):S22-S26.
7. Cleland VJ, Ball K, Magnussen C, Dwyer T, Venn A. Socioeconomic position and the tracking of physical activity and cardiorespiratory fitness from childhood to adulthood. *Am J Epidemiol*. 2009;170(9):1069-1077.
8. Andersen LB, Hasselstrom H, Gronfeldt V, Hansen SE, Karsten F. The relationship between physical fitness and clustered risk, and tracking of clustered risk from adolescence to young adulthood: eight years follow-up in the Danish Youth and Sport Study. *Int J Behav Nutr Phys Act*. 2004;1(1):6.
9. Ortega FB, Ruiz JR, Castillo MJ, et al. Low level of physical fitness in Spanish adolescents. Relevance for future cardiovascular health (AVENA study). *Rev Esp Cardiol*. 2005;58(8):898-909.
10. Ortega FB, Artero EG, Ruiz JR, et al. Physical fitness levels among European adolescents: the HELENA study. *Br J Sports Med*. 2011;45(1):20-29.
11. Research CES. *EUROFIT: Handbook for the EUROFIT Tests of Physical Fitness*. 2nd ed. Strasbourg, France: Committee of Experts on Sports Research; 1993.
12. Deporte IdCdeFye. *EUROFIT. TEST EUROPEO DE APTITUD FÍSICA*. Madrid, Spain: Ministerio de Educación y Ciencia; 1992.
13. Sauka M, Priedite IS, Artjuhova L, et al. Physical fitness in northern European youth: reference values from the Latvian Physical Health in Youth Study. *Scand J Public Health*. 2011;39(1):35-43.
14. EUROFIT. *Handbook for the EUROFIT Tests of Physical Fitness*. Rome, Italy: Council of Europe; 1988.
15. Meredith MD, Welk G. *FITNESSGRAM/ACTIVITYGRAM: Test Administration Manual*. 4th ed. Champaign, IL: Human Kinetics; 2007.
16. Meredith MD, Welk G. *FITNESSGRAM/ACTIVITYGRAM: Test Administration Manual*. 3rd ed. Champaign, IL: Human Kinetics; 2004.
17. Welk GJ, De Saint-Maurice Maduro PF, Laurson KR, Brown DD. Field evaluation of the new FITNESSGRAM(R) criterion-referenced standards. *Am J Prev Med*. 2011;41(4 suppl 2):S131-S142.
18. Welk GJ, Laurson KR, Eisenmann JC, Cureton KJ. Development of youth aerobic-capacity standards using receiver operating characteristic curves. *Am J Prev Med*. 2011;41(4 Suppl 2):S111-S116.
19. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med*. 1992;11(10):1305-1319.
20. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med*. 1998;17(4):407-429.
21. Chillon P, Ortega FB, Ferrando JA, Casajus JA. Physical fitness in rural and urban children and adolescents from Spain. *J Sci Med Sport*. 2011;14(5):417-423.
22. Woll A, Kurth BM, Oppen E, Worth A, Bos K. The 'Motorik-Modul' (MoMo): physical fitness and physical activity in German children and adolescents. *Eur J Pediatr*. 2011;170(9):1129-1142.
23. Sacchetti R, Ceciliani A, Garulli A, et al. Physical fitness of primary school children in relation to overweight prevalence and physical activity habits. *J Sports Sci*. 2012;30(7):633-640.
24. Volbekiene V, Gričiute A. Health-related physical fitness among schoolchildren in Lithuania: a comparison from 1992 to 2002. *Scand J Public Health*. 2007;35(3):235-242.
25. Catley MJ, Tomkinson GR. Normative health-related fitness values for children: analysis of 85347 test results on 9-17-year-old Australians since 1985. *Br J Sports Med*. 2013;47(2):98-108.
26. Armstrong ME, Lambert EV, Lambert MI. Physical fitness of South African primary school children, 6 to 13 years of age: discovery vitality health of the nation study. *Percept Mot Skills*. 2011;113(3):999-1016.

27. Malina RM, Bouchard C, Bar-Or O. *Growth, Maturation, and Physical Activity*. 2nd ed. Champaign, IL: Human Kinetics; 2004.
28. Haywood K, Getchell N. *Life Span Motor Development*. 3rd ed. Champaign, IL: Human Kinetics; 2001.
29. Davies PL, Rose JD. Motor skills of typically developing adolescents: awkwardness or improvement? *Phys Occup Ther Pediatr*. 2000;20(1):19-42.
30. Casajús JA, Ortega FB, Vicente-Rodríguez G, Leiva MT, Moreno LA, Ara I. Physical fitness, fat distribution and health in school-age children (7 to 12 years). *Rev Int Med Cienc Act Fís Deporte*. 2012;12(47):523-537.
31. Marta CC, Marinho DA, Barbosa TM, Izquierdo M, Marques MC. Physical fitness differences between prepubescent boys and girls. *J Strength Cond Res*. 2012;26(7):1756-1766.
32. Rowland TW. Evolution of maximal oxygen uptake in children. *Med Sport Sci*. 2007;50:200-209.
33. Armstrong N, Williams J, Balding J, Gentle P, Kirby B. The peak oxygen uptake of British children with reference to age, sex and sexual maturity. *Eur J Appl Physiol Occup Physiol*. 1991;62(5):369-375.
34. Martinez-Vizcaino V, Solera Martinez M, Notario Pacheco B, et al. Trends in excess of weight, underweight and adiposity among Spanish children from 2004 to 2010: the Cuenca study. *Public Health Nutr*. 2012;15(12):2170-2174.
35. Dencker M, Thorsson O, Karlsson MK, et al. Gender differences and determinants of aerobic fitness in children aged 8-11 years. *Eur J Appl Physiol*. 2007;99(1):19-26.
36. Alter MJ. *Science of Flexibility*. 3 ed. Champaign, IL: Human Kinetics; 2004.
37. Aznar S, Naylor PJ, Silva P, et al. Patterns of physical activity in Spanish children: a descriptive pilot study. *Child Care Health Dev*. 2011;37(3):322-328.
38. Dollman J, Maher C, Olds TS, Ridley K. Physical activity and screen time behaviour in metropolitan, regional and rural adolescents: a sectional study of Australians aged 9-16 years. *J Sci Med Sport*. 2012;15(1):32-37.
39. Haug E, Rasmussen M, Samdal O, et al. Overweight in school-aged children and its relationship with demographic and lifestyle factors: results from the WHO-Collaborative Health Behaviour in School-aged Children (HBSC) study. *Int J Public Health*. 2009;54(suppl 2):167-179.
40. Hasselstrom H, Hansen SE, Froberg K, Andersen LB. Physical fitness and physical activity during adolescence as predictors of cardiovascular disease risk in young adulthood. Danish Youth and Sports Study An Eight-year Follow-up Study. *Int J Sports Med*. 2002;23(suppl 1):S27-S31.
41. Twisk JW, Kemper HC, van Mechelen W. Prediction of cardiovascular disease risk factors later in life by physical activity and physical fitness in youth: general comments and conclusions. *Int J Sports Med*. 2002;23(suppl 1):S44-S49.
42. Mota J, Flores L, Ribeiro JC, Santos MP. Relationship of single measures of cardiorespiratory fitness and obesity in young schoolchildren. *Am J Hum Biol*. 2006;18(3):335-341.
43. Ruiz JR, Ortega FB, Rizzo NS, et al. High cardiovascular fitness is associated with low metabolic risk score in children: the European Youth Heart Study. *Pediatr Res*. 2007;61(3):350-355.
44. Ortega FB, Ruiz JR, Sjostrom M. Physical activity, overweight and central adiposity in Swedish children and adolescents: the European Youth Heart Study. *Int J Behav Nutr Phys Act*. 2007;4:61.
45. Ekelund U, Sardinha LB, Anderssen SA, et al. Associations between objectively assessed physical activity and indicators of body fatness in 9- to 10-y-old European children: a population-based study from 4 distinct regions in Europe (the European Youth Heart Study). *Am J Clin Nutr*. 2004;80(3):584-590.
46. Meredith MD, Welk G. *Cooper Institute (Dallas Tex.). FITNESSGRAM & ACTIVITYGRAM: Test Administration Manual*. 4th ed. Champaign, IL: Human Kinetics; 2010.
47. Castro-Pineiro J, Ortega FB, Keating XD, Gonzalez-Montesinos JL, Sjostrom M, Ruiz JR. Percentile values for aerobic performance running/walking field tests in children aged 6 to 17 years: influence of weight status. *Nutr Hosp*. 2011;26(3):572-578.
48. Castro-Pinero J, Gonzalez-Montesinos JL, Mora J, et al. Percentile values for muscular strength field tests in children aged 6 to 17 years: influence of weight status. *J Strength Cond Res*. 2009;23(8):2295-2310.
49. Anderssen SA, Cooper AR, Riddoch C, et al. Low cardiorespiratory fitness is a strong predictor for clustering of cardiovascular disease risk factors in children independent of country, age and sex. *Eur J Cardiovasc Prev Rehabil*. 2007;14(4):526-531.
50. Armstrong N, Tomkinson G, Ekelund U. Aerobic fitness and its relationship to sport, exercise training and habitual physical activity during youth. *Br J Sports Med*. 2011;45(11):849-858.
51. Martinez Vizcaino V, Salcedo Aguilar F, Franquelo Gutierrez R, et al. Assessment of an after-school physical activity program to prevent obesity among 9- to 10-year-old children: a cluster randomized trial. *Int J Obes (Lond)*. 2008;32(1):12-22.
52. Moya Martinez P, Sanchez Lopez M, Lopez Bastida J, et al. Cost-effectiveness of an intervention to reduce overweight and obesity in 9-10-year-olds. The Cuenca study. *Gac Sanit*. 2011;25(3):198-204.
53. Chillon P, Ortega FB, Ruiz JR, et al. Active commuting to school in children and adolescents: an opportunity to increase physical activity and fitness. *Scand J Public Health*. 2010;38(8):873-879.



OPEN

Secular trends in health-related physical fitness among 11–14-year-old Croatian children and adolescents from 1999 to 2014

Mario Kasović^{1,2}, Lovro Štefan^{1,2✉} & Vilko Petrić³

The main purpose of the study was to analyze secular trends of health-related physical fitness in 7–14-year-old Croatian children and adolescents from 1999 and 2014. In this observational cross-sectional study, we recruited 5077 children and adolescents between ages 11 and 14 (50.8% girls) from five primary schools located in the capital city of Zagreb. Physical fitness performance was tested from 1999 until 2014. Physical fitness performance included: (1) body-mass index (measure of body size), (2) standing broad jump (measure of lower-body power), (3) polygon backwards (measure of general coordination and agility), (4) sit-ups in 60 s (measure of upper-body strength), (5) sit-and-reach test (measure of flexibility) and (6) 6-min run test (measure of cardiorespiratory fitness). Boys performed better in all physical fitness tests, except for sit-and-reach test ($p < 0.001$). In boys, between 1999 and 2014, body size, upper-body strength and coordination/agility increased, while flexibility, lower-body power and cardiorespiratory fitness decreased. During the same period, girls experienced an increase in body size, lower-body power, upper-body strength, coordination/agility and flexibility, while cardiorespiratory fitness decreased. This study shows that cardiorespiratory fitness, flexibility and coordination/agility decrease, while upper-body strength increases in both sexes. These findings should serve as an avenue for national monitoring system to screen and track biological development in children and adolescents.

Physical fitness is an important and one of the most relevant indicators of health and sports participation in young populations^{1–3}. Evidence suggests that higher levels of physical fitness in childhood may prevent from developing some communicable and non-communicable diseases in adulthood^{2,4,5}. Apart from general physical fitness, separate components of physical fitness (body composition, cardiorespiratory, muscular and motor) have been inversely associated with cardiovascular and metabolic diseases, stroke and all-cause mortality^{6–8}.

Despite the fact that risk factors for diseases usually manifest during adulthood, studies have shown that childhood and adolescence are critical time-points for maintaining ‘good’ lifestyle habits and develop high physical fitness levels⁹ due to well-tracking characteristics of physical fitness from childhood to adulthood¹. However, a common perception has been embraced that the physical fitness in youth has declined because of more time spent in sedentary behavior, lack of physical activity and consuming fat-rich food¹⁰. Such decrease is, however not uniform in different countries and physical fitness modalities^{11,12}.

Secular trends of physical fitness in children and adolescents have been well-studied in the past decades^{10,12–35}. In general, values in cardiorespiratory, muscular and flexibility fitness decline, and values in body-mass index (as a measure of body size) increase, especially in work that goes back further in time (between 1960 and 1980s), while there have been conflicting findings for motor fitness in terms of a decrease, increase or stagnation³⁶. Such inconsistencies may depend on the sample size, age- and sex-specific populations studied, and physical fitness tests applied¹².

¹Department of General and Applied Kinesiology, Faculty of Kinesiology, University of Zagreb, Horvaćanski zavoj 15, Zagreb, Croatia. ²Department of Sport Motorics and Methodology in Kinesiology, Faculty of Sports Studies, Masaryk University, Brno, Czech Republic. ³Department of Educational Studies, Education, University of Rijeka, Rijeka, Croatia. ✉email: lovro.stefan1510@gmail.com

In recent years, several European countries have proposed national monitoring systems for screening and monitoring biological development of children and adolescents^{10,37–39}. Although such measures through projects (like the HBSC study) have been implemented and tested in Croatian youth, the most recent study has shown that $\approx 50\%$ of youth in Croatia do not meet the recommended levels of physical activity, $\approx 25\%$ exceed 2 h of screen time per day and have high blood pressure, and $\approx 20\%$ have excess weight, leading to overweight and obesity status⁴⁰. Thus, being able to keep tracking separate components of physical fitness would benefit from establishing national-wide policies and strategies aiming to standardize the assessment of children's somatic development and functional capacity in critical periods of childhood and adolescence¹⁰.

Therefore, the main purpose of the study was to analyze secular trends of health-related physical fitness in 7–14-year-old Croatian children and adolescents from 1999 and 2014. Based on previous evidence, we hypothesized that poorer values would be observed for body size (an increase), cardiorespiratory fitness and flexibility (a decrease), with improvements in muscular fitness.

Methods

Study participants. In this observational cross-sectional study, we recruited 5,077 children and adolescents between ages 11 and 14 (50.8% girls) from five primary schools located in the capital city of Zagreb. To be included, students had to be healthy and participated in physical education classes at the time of study and had to be not specifically trained for performance in the administered tests¹². Physical fitness performance was tested from 1999 until 2014 in different sex- and age-specific samples of children and adolescents. The study was approved by the Faculty of Kinesiology, University of Zagreb, Croatia 8 ethical code number: 02/2014). Parent of each participant and all participants gave informed written consent before enrollment into the study for participation and approving to publish the information in an online open access publication. Analyses and procedures performed in the study were conducted in accordance with the Declaration of Helsinki⁴¹. The tests were performed each year in September.

Health-related physical fitness. Height and weight were objectively measured in light clothing by using stadiometer and digital scale with a precision of 0.1 cm and 0.1 kg. *Body-mass index* was calculated by dividing weight in kg with height in m^2 [weight(kg)/height(m)²]. *Standing broad jump* tests jumping distance from a standing start ('frog leap'). While performing the jumps, each child was asked to bend their knees with their arms in front of them, parallel to the ground, then to swing both arms, push off vigorously and jump forward as far as possible, trying to land with their feet together and stay upright⁴². *Polygon backwards* tests motor coordination and agility, where the subject had to crawl backwards and cover a 10-m distance. The starting position for the children was behind the starting line as the children faced backwards on all fours. The tester gave the starting signal. The test included backward crawling (A) over and (B) under the 35 cm high obstacles that were placed at (A) 3 m and (B) 6 m from the starting line. The task was measured in tenths of a second (0.1 s)⁴³. *Sit-up test* evaluates abdominal muscles function as number of sit-ups completed from lying position (knees bent at a 90°) in 60 s. Children were seated on the floor, backs straight, hands clasped behind their neck, knees bent at 90° with heels and feet flat on the mat. They then lay down on their backs, shoulders touching the mat, and returned to the sitting position with their elbows out in front to touch their knees, keeping the hands clasped behind their neck the whole time⁴². *Sit-and reach test* assesses the level of flexibility. By sitting on the floor or a mat, legs straight under the angle of 90°, the person being tested reached forward with the arms (hands overlapping). The distance of reach was measured in centimeters using a measuring non-elastic tape attached on the floor⁴⁴. *6-min run test* is used to assess the level of cardiorespiratory fitness. Participants were asked to run or walk constantly for 6 min. The distance covered by each participant was measured by test leaders²⁴.

Data analysis. Basic descriptive statistics are presented as mean and standard deviation (SD). Kolmogorov–Smirnov test was applied to identify outliers, which were subsequently excluded. Analysis of variance (ANOVA) was used to determine the differences between sexes in all physical fitness components. To assess the associations between the assessment year and each physical fitness aspect, we used generalized estimating equations. The working correlation matrix was set to exchangeable in all analyses. The regression models were tested for several assumptions: (1) multicollinearity diagnostics using variance inflation index, (2) normality of residuals using the normal probability plot and histogram of residuals, and (3) heteroscedasticity using the standardized residuals vs. predicted plot. All assumptions were met for all regression models. Cohen's *d* effect size calculations were performed to compare the magnitude of difference between 1999 and 2014, with thresholds set at 0.2, 0.21–0.5, 0.51–0.8, and 0.8 for trivial, small, moderate, and large magnitudes of effect, respectively. Sex-specific analyses were performed, since there were significant differences between boys and girls in all physical fitness tests ($p < 0.05$). Two-sided *p*-values were used, and significance was set at $\alpha < 0.05$. All the analyses were calculated in Statistical Packages for Social Sciences v.23 (SPSS, Chicago, IL, United States).

Ethics approval and consent to participate. The study was approved by the Faculty of Kinesiology, University of Zagreb, Croatia. The informed consent voluntarily was signed by the participants, participants' parents or their guardians.

Consent for publication. The informed consent voluntarily was signed by the participants, participants' parents or their guardians.

Sex/year	Health-related physical fitness components					
	Body-mass index (kg/m ²)	Standing broad jump (cm)	Polygon backwards (sec)	Sit-ups in 60 s (#)	Sit-and-reach test (cm)	6-min run test (m)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Boys						
1999	18.7 (2.7)	175.6 (26.7)	12.2 (3.9)	35.1 (7.9)	51.0 (10.6)	1173.7 (194.5)
2000	18.6 (2.8)	172.3 (24.8)	13.8 (4.5)	38.7 (8.3)	57.2 (14.1)	1168.5 (174.0)
2001	18.8 (2.4)	172.1 (25.4)	13.5 (4.2)	38.7 (7.4)	55.2 (13.8)	1142.4 (168.3)
2002	18.9 (2.6)	171.2 (20.8)	13.3 (3.8)	38.9 (6.9)	54.4 (9.9)	1109.3 (123.9)
2003	19.1 (2.9)	175.8 (21.5)	12.5 (3.6)	39.4 (7.9)	62.0 (10.0)	1175.4 (172.3)
2004	19.5 (3.4)	171.8 (23.5)	13.0 (4.0)	41.2 (8.8)	64.4 (11.8)	1179.1 (170.2)
2005	19.9 (4.0)	175.5 (25.6)	14.6 (4.4)	38.2 (8.4)	63.5 (11.8)	1144.6 (191.3)
2006	20.2 (4.0)	171.8 (27.3)	15.0 (4.6)	37.4 (8.5)	54.8 (9.7)	1168.5 (194.2)
2007	20.1 (4.2)	169.3 (27.3)	15.5 (5.1)	36.8 (7.5)	56.6 (11.1)	1144.4 (199.9)
2008	20.0 (3.8)	165.6 (24.7)	15.4 (4.5)	37.0 (7.3)	55.4 (11.5)	1138.7 (178.2)
2009	19.8 (3.7)	170.1 (25.1)	14.5 (4.3)	38.7 (8.1)	61.1 (14.3)	1085.3 (220.1)
2010	19.8 (3.2)	170.1 (29.1)	14.2 (3.5)	39.2 (8.4)	53.6 (9.2)	1164.9 (187.9)
2011	20.1 (3.3)	172.7 (27.1)	15.5 (4.8)	40.1 (8.6)	58.7 (11.5)	1131.5 (221.8)
2012	19.9 (3.3)	175.7 (26.3)	13.0 (3.1)	41.2 (8.2)	57.2 (11.4)	909.3 (120.5)
2013	20.1 (3.2)	174.6 (29.8)	11.1 (1.1)	41.2 (8.5)	44.3 (9.6)	846.9 (110.8)
2014	20.1 (3.5)	169.9 (21.9)	10.9 (1.1)	40.7 (8.1)	44.7 (9.2)	828.3 (140.7)
Total	19.6 (3.5)	172.3 (25.6)	13.7 (4.2)	38.9 (8.3)	56.4 (12.6)	1029.2 (175.9)
Girls						
1999	19.0 (3.1)	156.0 (23.8)	17.0 (5.8)	33.3 (8.9)	57.0 (11.4)	1108.5 (220.3)
2000	18.3 (2.5)	163.1 (21.1)	14.5 (3.5)	33.4 (7.5)	66.5 (14.9)	1014.8 (159.1)
2001	19.3 (3.3)	148.3 (24.2)	22.4 (9.4)	31.3 (10.2)	55.7 (11.9)	1087.4 (276.0)
2002	18.6 (2.5)	159.4 (18.9)	16.4 (4.7)	33.4 (5.1)	61.2 (9.9)	986.9 (103.2)
2003	19.3 (3.0)	160.7 (21.6)	14.9 (4.5)	34.0 (7.2)	71.5 (11.2)	1017.3 (149.7)
2004	19.5 (4.0)	160.9 (23.1)	14.8 (5.5)	36.1 (8.7)	72.6 (11.9)	1032.0 (144.4)
2005	19.2 (3.3)	161.2 (20.7)	15.7 (5.7)	36.8 (7.9)	72.6 (11.7)	1070.7 (134.7)
2006	19.0 (3.2)	159.8 (22.1)	16.4 (5.5)	35.8 (7.6)	63.9 (11.0)	1067.6 (140.1)
2007	19.0 (2.9)	160.0 (23.5)	17.2 (4.5)	34.9 (7.4)	63.3 (13.5)	1040.3 (133.3)
2008	18.8 (2.8)	152.0 (22.3)	18.2 (5.1)	33.0 (7.4)	59.5 (11.0)	1022.0 (133.9)
2009	19.1 (2.9)	154.4 (21.4)	16.5 (4.5)	35.2 (5.9)	65.7 (12.5)	942.0 (164.5)
2010	18.7 (2.4)	160.8 (21.8)	15.6 (4.6)	36.1 (6.8)	67.8 (12.6)	1017.3 (186.8)
2011	18.9 (2.6)	163.9 (22.0)	16.2 (4.2)	37.5 (6.3)	67.0 (11.8)	1038.2 (160.5)
2012	18.9 (3.1)	161.5 (21.4)	14.3 (3.8)	37.2 (6.3)	64.6 (11.7)	848.3 (148.5)
2013	19.0 (3.0)	160.6 (23.5)	11.6 (0.9)	38.2 (6.4)	51.7 (9.4)	797.9 (130.5)
2014	19.1 (2.9)	161.2 (20.4)	11.6 (1.0)	39.3 (7.0)	52.0 (9.8)	680.5 (128.7)
Total	19.0 (3.0)	158.4 (22.6)	16.3 (5.9)	34.9 (8.0)	63.2 (13.6)	973.1 (156.1)

Table 1. Basic descriptive statistics of the study participants ($N = 5077$).

Results

Basic descriptive statistics are presented in Table 1. Boys had higher body-mass index value ($p < 0.001$) and performed better in standing broad jump ($p < 0.001$), polygon backwards ($p < 0.001$), sit-ups in 60 s ($p < 0.001$) and 6-min run test ($p < 0.001$), compared to girls. Girls scored better in sit-and-reach test ($p < 0.001$).

Secular trends in health-related physical fitness are presented in Fig. 1. From 1999 to 2014, body-mass index and the number of sit-ups in 60 s increased by 7.5% (standardized $\beta = 0.12$, $p < 0.001$, ES = 0.45) and 15.9% (standardized $\beta = 0.13$, $p < 0.001$, ES = 0.70), while the results in standing broad jump, polygon backwards, sit-and-reach test and 6-min run test decreased by 3.2% (standardized $\beta = -0.02$, $p = 0.327$, ES = 0.23), 10.7% (standardized $\beta = -0.01$, $p = 0.939$, ES = 0.15), 12.4% (standardized $\beta = -0.14$, $p < 0.001$, ES = 0.61) and 12.3% (standardized $\beta = -0.55$, $p < 0.001$, ES = 2.00) in boys, respectively. During the same period, girls experienced a significant increase in standing broad jump and sit-ups in 60 s by 3.3% (standardized $\beta = 0.07$, $p = 0.004$, ES = 0.23) and 18.0% (standardized $\beta = 0.21$, $p < 0.001$, ES = 0.75), while the results in polygon backwards, sit-and-reach test and 6-min run test decreased by 31.8% (standardized $\beta = -0.20$, $p < 0.001$, ES = 1.30), 8.8% (standardized $\beta = 0.03$, $p = 0.268$, ES = 0.47) and 38.6% (standardized $\beta = -0.52$, $p < 0.001$, ES = 2.37), respectively.

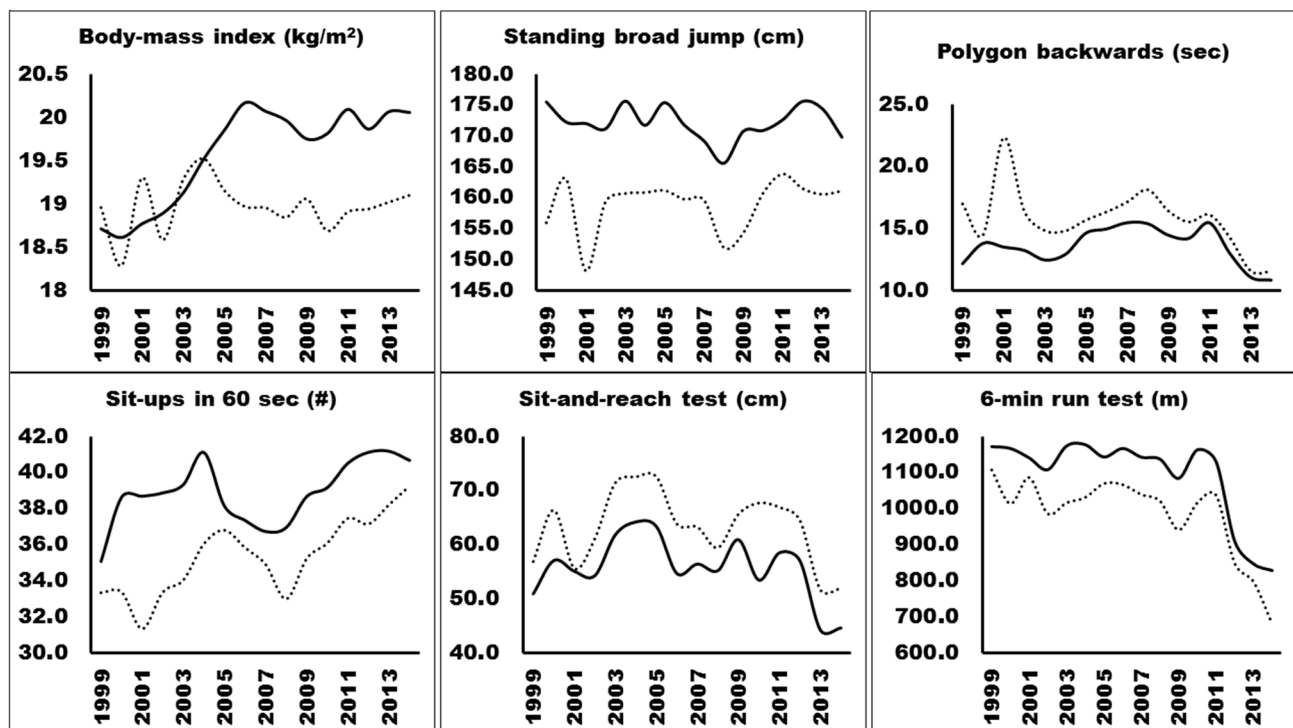


Figure 1. Secular trends of in health-related physical fitness among 11 to 14-year-old boys (black full line) and girls (discontinuous line) from 1999 to 2014.

Discussion

The main purpose of the study was to analyze secular trends of health-related physical fitness in 7–14-year-old Croatian children and adolescents from 1999 and 2014. The main findings are: (1) boys tend to perform better in upper-body strength tests, yet decrease the results in flexibility and cardiorespiratory fitness tests, (2) girls tend to perform better in upper-body strength, lower-body power and coordination/agility tests, yet decrease the results in cardiorespiratory fitness test.

Previous evidence has shown that cardiorespiratory performance declines in children and adolescents^{12,16–19}. Our results confirm a negative trend of aerobic fitness in both sexes, which is alarming because low cardiorespiratory fitness is associated with development of numerous diseases and all-cause mortality^{1–9}. It has been well-documented, that cardiorespiratory fitness correlates with physical activity⁴⁵. In Croatia, studies have shown a negative trend of meeting the recommended levels of physical activity from 2002 to 2010⁴⁶, possibly causing a significant decline in cardiorespiratory fitness. Such findings have been obtained previously¹², pointing out that by enhancing cardiorespiratory fitness, one would be expecting to increase the level of physical activity and vice versa. Of note, we observed that the level of cardiorespiratory fitness had declined from 1999 to 2014 by $\approx 22.0\%$ for both sexes, which is similar to previous studies using 20-m shuttle run test as a proxy of cardiorespiratory fitness¹⁶. In general, a recent systematic review aiming to report trends in physical fitness has shown, that the declining trend in cardiorespiratory fitness is more obvious in studies with more time periods or longer follow-ups³⁶, which has also been supported by large-scale epidemiological findings¹⁸.

The observation of a declining trend of flexibility in this study is consistent to previous studies conducted among similar age groups^{12,21,23,32,36,47}. Flexibility performance is associated to body size (mainly to body-mass index), and since we observed a significant increase of body-mass index from 1999 to 2014, it is possible that such increase would correlate to a negative trend of flexibility²¹. Indeed, there has been a general consensus that flexibility in children and adolescents steadily declines, which has been proven previously³⁶. Being able to support such statement comes from similar methodological testing procedure and using sit-and-reach test, as a proxy of flexibility.

In both sexes, we observed a positive trend of upper-body strength (by using the number of sit-ups in 30 s), which is consistent to previous findings^{11,12,23,48,49}. On the other hand, lower-body power had been stagnating between 1999 and 2014, while in girls a significant positive trend was observed^{48,49}. Specifically, studies by Albon et al.⁴⁸ and Huotari et al.⁴⁹ showed no significant changes for standing broad jump performance. On the other hand, a large-scale systematic review conducted among 27 countries and 6–19-year-old children and adolescents has shown that power ability had been slightly improving between 1958 and 2003, regardless of gender, age or socioeconomic status¹¹. The discrepancy between this and previous studies may be explained by using different sample sizes, age-groups, statistical analyses and time-periods for assessing secular trends. According to recent systematic review³⁶, the majority of the primary studies reported a decline in strength performance. Although we found a significant positive change of standing broad jump performance in girls, the effect size between 1999 and 2014 was only trivial.

Coordination/agility performance assessed through polygon backwards had been improving in both sexes, with significantly greater changes in girls. Such findings are similar to previous studies^{12,24}. However, a study by Eberhardt et al.³⁶ showed that the same number of studies reported a decrease, increase or stagnation. The inconsistency between the studies comes from the construct of coordinative abilities, which are being complex and multidimensional³⁶. Second, different coordination tasks can measure different aspects of coordination³⁶. Therefore, it is difficult to compare the findings in this measure, since different motor tests to assess coordination/agility (obstacle course, balancing backwards, target throwing or one-leg stand) have been used previously³⁶. Nevertheless, polygon backwards is a reliable and valid test to assess the level of coordination/agility in children⁴².

Finally, an increased trend in body-mass index of this study is consistent with findings from previous studies^{12,13,21,50}. The ecological shift towards not meeting the recommended levels of physical activity, exceeding 2 h of screen time per day and having high blood pressure and excessive weight in Croatian youth⁴⁰ may have led to a negative trend of body-mass index, particularly in boys. However, stronger trend in boys observed in this study may be attributed to different body composition levels between sexes, where boys have higher values of lean-muscle mass and lower values of fat-mass, compared to girls. Unfortunately, body-mass index cannot discriminate between lean-muscle mass and fat mass and, therefore, changes in body size need to be interpreted with caution.

This study has several limitations. First, by using a cross-sectional design, we cannot determine the causality of the association. Second, we used the 6-min run test as a proxy of cardiorespiratory fitness. By using more objective methods, like the treadmill test, the findings might have been different. Also, the time needed for polygon backwards was measured manually and not with the use of a light barrier. Finally, each test in a specific year was performed only once, limiting maximal efficiency and performance. Therefore, future studies should screen and monitor the level of physical fitness during the follow-up period by using standardized tests performed a few times, in order to determine separate and overall physical fitness trends.

The study confirms the hypothesis of previous evidence, that cardiorespiratory fitness has negative secular trend, yet upper-body strength tends to have positive secular trend in both sexes. Boys and girls seem to produce different secular trends in other physical fitness components, pointing out that sex-specific surveillance system of physical fitness should be incorporated within the school system, in order to monitor and track separate and overall physical fitness in children and adolescents.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 2 November 2020; Accepted: 17 May 2021

Published online: 26 May 2021

References

- Ortega, F. B., Ruiz, J. R., Castillo, M. J. & Sjörström, M. Physical fitness in childhood and adolescence: A powerful marker of health. *Int. J. Obes. (Lond)* **32**, 1–11 (2007).
- Hamer, M., O'Donovan, G., Batty, G. D. & Stamatakis, E. Estimated cardiorespiratory fitness in childhood and cardiometabolic health in adulthood: 1970 British Cohort Study. *Scand. J. Med. Sci. Sports* **30**, 923–938 (2020).
- Ruiz, J. R. et al. Predictive validity of health-related fitness in youth: A systematic review. *Br. J. Sports Med.* **43**, 99–932 (2009).
- Antero-Jacquemin, J. et al. The heart of the matter: Years-saved from cardiovascular and cancer deaths in an elite athlete cohort with over a century of follow-up. *Eur. J. Epidemiol.* **33**, 531–543 (2018).
- Myers, J. et al. Exercise capacity and mortality among men referred for exercise testing. *N. Engl. J. Med.* **346**, 793–801 (2002).
- Timpka, S. et al. Muscle strength in adolescent men and risk of cardiovascular disease events and mortality in middle age: A prospective cohort study. *BMC Med.* **12**, 62 (2014).
- Lee, C. D., Blair, S. N. & Jackson, A. S. Cardiorespiratory fitness, body composition, and all-cause and cardiovascular disease mortality in men. *Am. J. Clin. Nutr.* **69**, 373–380 (1999).
- Kuwahara, K. et al. Association of cardiorespiratory fitness and overweight with risk of type 2 diabetes in Japanese men. *PLoS One* **9**, e98508 (2014).
- Högström, G. et al. Risk factors assessed in adolescence and the later risk of stroke in men: A 33-year follow-up study. *Cerebrovasc. Dis.* **39**, 63–71 (2015).
- Potočnik, ŽL., Jurak, G. & Starc, G. Secular trends of physical fitness in twenty-five birth cohorts of Slovenian children: A population-based study. *Front. Public Health* **8**, 561273 (2020).
- Tomkinson, G. R., Olds, T. S. & Borms, J. Who are the Eurofittest?. *Med. Sport. Sci.* **50**, 104–128 (2007).
- Vencunas, T. et al. Secular trends in physical fitness and body size in Lithuanian children and adolescents between 1992 and 2012. *J. Epidemiol. Community Health* **71**, 181–187 (2017).
- Malina, R. M. Secular trends in growth, maturation and physical performance: A review. *Anthropol. Rev.* **67**, 3–31 (2004).
- Wedderkopp, N. et al. Secular trends in physical fitness and obesity in Danish 9-year-old girls and boys: Odense School Child Study and Danish substudy of the European Youth Heart Study. *Scand. J. Med. Sci. Sports* **14**, 150–155 (2004).
- Malina, R. M. Physical fitness of children and adolescents in the United States: Status and secular change. *Med. Sport. Sci.* **50**, 67–90 (2007).
- Møller, N. C. et al. Secular trends in cardiorespiratory fitness and body mass index in Danish children: The European Youth Heart Study. *Scand. J. Med. Sci. Sports* **17**, 331–339 (2007).
- Stratton, G. et al. Cardiorespiratory fitness and body mass index of 9–11-year-old English children: A serial cross-sectional study from 1998 to 2004. *Int. J. Obes.* **31**, 1172–1178 (2007).
- Tomkinson, G. R. & Olds, T. S. Secular changes in pediatric aerobic fitness test performance: The global picture. *Med. Sport. Sci.* **50**, 46–66 (2007).
- Sandercock, G. et al. Ten-year secular declines in the cardiorespiratory fitness of affluent English children are largely independent of changes in body mass index. *Arch. Dis. Child.* **95**, 46–57 (2010).
- Runhaar, J. et al. Motor fitness in Dutch youth: Differences over a 26-year period (1980–2006). *J. Sci. Med. Sport.* **13**, 323–328 (2010).

21. Dos Santos, F. K. *et al.* Secular trends in physical fitness of Mozambican school-aged children and adolescents. *Am. J. Hum. Biol.* **27**, 206–206 (2015).
22. Tyler, R. *et al.* Ten-year secular changes in selected health and fitness parameters of 10–11 years old Swansea school children-2003–2013. *Adv. Obes. Weight Manag. Control* **3**, 267–271 (2015).
23. Costa, A. M. *et al.* Secular trends in anthropometrics and physical fitness of young Portuguese school-aged children. *Acta Med. Portuguesa* **30**, 108–114 (2017).
24. Spengler, S. *et al.* Trends in motor performance of first graders: A comparison of cohorts from 2006 to 2015. *Front. Pediatr.* **5**, 206 (2017).
25. Aaberge, K. & Mamen, A. A comparative study of fitness levels among Norwegian youth in 1988 and 2001. *Sports* **7**, 50 (2019).
26. Sandercock, G. R. & Cohen, D. D. Temporal trends in muscular fitness of English 10-year-olds 1998–2014: An allometric approach. *J. Sci. Med. Sport.* **22**, 201–205 (2019).
27. Kocić, J. *et al.* Muscle strength test performance changes over time in Serbian children. *Acta Med. Medianae* **58**, 154–160 (2019).
28. Fraser, B. J. *et al.* The great leap backward: Changes in the jumping performance of Australian children aged 11–12-years between 1985 and 2015. *J. Sports Sci.* **37**, 748–754 (2019).
29. Shigaki, G. B. *et al.* Secular trend of physical fitness indicators related to health in children. *J. Hum. Growth Dev.* **29**, 381–389 (2019).
30. Gaya, A. R. *et al.* Temporal trends in physical fitness and obesity among Brazilian children and adolescents between 2008 and 2014. *J. Hum. Sport. Exerc.* **15**, 549–558 (2019).
31. Dong, Y. *et al.* Trends in physical fitness, growth, and nutritional status of Chinese children and adolescents: A retrospective analysis of 1.5 million students from six successive national surveys between 1985 and 2014. *Lancet Child. Adolesc. Health* **3**, 871–880 (2019).
32. Colley, R. C. *et al.* Trends in physical fitness among Canadian children and youth. *Health Rep.* **30**, 3–13 (2019).
33. Bi, C., Zhang, F., Gu, Y., Song, Y. & Cai, X. Secular trend in the physical fitness of Xinjiang children and adolescents between 1985 and 2014. *Int. J. Environ. Res. Public Health.* **17**, 2195 (2020).
34. Hanssen-Doose, A. *et al.* Population-based trends in physical fitness of children and adolescents in Germany, 2003–2017. *Eur. J. Sport Sci.* <https://doi.org/10.1080/17461391.2020.1793003> (2020) (Epub ahead of print).
35. Pinoniemi, B. K. *et al.* Temporal trends in the standing broad jump performance of United States children and adolescents. *Res. Q. Exerc. Sport.* <https://doi.org/10.1080/02701367.2019.1710446> (2020).
36. Eberhardt, T. *et al.* Secular trends in physical fitness of children and adolescents: A review of large-scale epidemiological studies published after 2006. *Int. J. Environ. Res. Public Health* **17**, 5671 (2020).
37. Csányi, T. *et al.* Overview of the Hungarian national youth fitness study. *Res. Q. Exerc. Sport.* **86**, 3–12 (2015).
38. Henriques-Neto, D. *et al.* Test-retest reliability of physical fitness tests among young athletes: the FITescola battery. *Clin. Physiol. Funct. Imaging* **40**, 173–182 (2020).
39. Joensuu, L. *et al.* Objectively measured physical activity, body composition and physical fitness: Cross-sectional associations in 9- to 15-year-old children. *Eur. J. Sport. Sci.* **18**, 882–892 (2018).
40. Radman, I., Sorić, M. & Mišigoj-Duraković, M. Prevalence of key modifiable cardiovascular risk factors among urban adolescents: The CRO-PALS study. *Int. J. Environ. Res. Public Health* **17**, 3162 (2020).
41. World Health Association. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bull. World Health Organ.* **79**, 373–374 (2001).
42. Štefan, L., Paradžik, P. & Sporiš, G. Sex and age correlations of reported and estimated physical fitness in adolescents. *PLoS One* **14**, e0219217 (2019).
43. Sekulić, D. *et al.* Judo training is more effective for fitness development. *Pediatr. Exerc. Sci.* **18**, 329–338 (2006).
44. PCPFS (President's Council on Physical Fitness and Sports). The president's challenge physical fitness test: V-sit reach. 2012. <https://www.presidentschallenge.org/challenge/physical/activities/v-sit-reach.shtml>. Assessed 29 Oct 2020.
45. Dencker, M. *et al.* Daily physical activity and its relation to aerobic fitness in children aged 8–11 years. *Eur. J. Appl. Physiol.* **96**, 587–592 (2006).
46. Kalman, M. *et al.* Secular trends in moderate-to-vigorous physical activity in 32 countries from 2002 to 2010: A cross-national perspective. *Eur. J. Public Health* **25**, 37–40 (2015).
47. Vollbekiene, V. & Gričiūtė, A. Health-related physical fitness among schoolchildren in Lithuania: A comparison from 1992 to 2002. *Scand. J. Public Health* **35**, 235–242 (2007).
48. Albion, H. M., Hamlin, M. J. & Ross, J. J. Secular trends and distributional changes in health and fitness performance variables of 10–14-year-old children in New Zealand between 1991 and 2003. *Br. J. Sports Med.* **44**, 263–269 (2010).
49. Huotari, P. R. T. *et al.* Secular trends in muscular fitness among Finnish adolescents. *Scand. J. Public Health* **38**, 739–747 (2010).
50. Dollman, J. & Olds, T. S. Secular changes in fatness and fat distribution in Australian children matched for body size. *Int. J. Pediatr. Obes.* **1**, 109–113 (2006).

Acknowledgements

The authors would like to thank all the physical education teachers, children and their parents or guardians for enthusiastic participation in the study.

Author contributions

V.P. conducted the study. L.Š. analyzed the data. M.K., L.Š. and V.P. wrote and drafted the manuscript. The author(s) read and approved the final version of the manuscript.

Funding

The paper was self-funded.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.Š.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Sexual Dimorphism in Body Composition, Weight Status and Growth in Prepubertal School Children from Rural Areas of Eastern Austria

S. Kirchengast and V. Steiner

Institute for Anthropology, University of Vienna, Vienna, Austria

ABSTRACT

Sexual dimorphism in stature, weight status and body composition were analyzed in a sample of 398 prepubertal children (213 girls, 185 boys) ageing between 7 and 10. Furthermore the prevalence of overweight was tested. Body composition parameters were determined using TBF 105 Body composition analyzer according to BIA-method. Highly significant sex differences in body composition were observed ($p < 0.001$). In contrast, stature, weight and BMI showed no significant differences between the two sexes. Nevertheless, a significant higher portion ($p < 0.05$) of girls (29%) corresponded to the definition overweight according to ASNS (Austrian Survey of Nutritional Status), while only 20% of the boys fell into the category overweight. The results of the present study showed not only significant sex differences in body composition, especially in fat mass, long before puberty onset, but also a significantly higher prevalence of overweight among prepubertal girls in comparison to prepubertal boys.

Introduction

The health status of individuals or populations is closely linked to nutritional status. While malnutrition and underweight are typical nutritional problems of Third World countries, overweight and obesity belong to the most important health risk factors of contemporary in-

dustrialized societies¹. Obesity in association with a centralized fat distribution is associated with an increased risk for the development of the so-called metabolic syndrome and is also described as an important risk factor for colon and breast cancer². However, obesity, especially centralized obesity, is not only a health problem of adults, especially middle-aged people,

Received for publication January 28, 2001.

DEFENDANT'S
EXHIBIT

14

21

who represent the main risk group to develop a metabolic syndrome and colon or breast cancer, obesity and overweight are also marked problems during childhood and adolescence¹. Especially during the last three decades the prevalence of overweight and obesity in childhood and adolescents is rising^{1,3–5}. This secular trend may be the result of behavioral changes taking place during the past three decades such as steadily decreased energy expenditure, low physical activity levels, spending a lot time of watching TV^{6–9}, and changed nutritional habits¹⁰. Unfortunately obesity present in childhood or adolescence increases the likelihood of adult obesity and therefore morbidity and mortality during adult phase of life^{11,12}. Especially a centralized fat distribution which is known to be associated with the health risks mentioned above during adulthood, seems to persist throughout childhood and adolescence up to adult age¹³. A centralized fat distribution and a high amount of fat tissue are typical for overweight and obese children and adults. However body composition and fat distribution vary with age and sex. During childhood fat distribution patterns and body composition seems to be similar in girls and boys, gender differences occur during adolescence^{14,15}. This is true of fat distribution patterns as well as body composition characteristics. Boys accumulate relative more fat on the upper body, while girls appear to gain fat tissue on upper and lower body at a similar pace. In general, gender differences in body composition especially in the body fat content are usually said to arise during adolescence, when the energetic demands of reproduction make additional fat deposits in females as energy stores necessary. Gender differences in body composition during prepubertal phase of life are considered to be slight¹⁶ and boys and girls exhibit similar average weights and heights¹⁷. The application of dual energy X-rays absorptiometry (DEXA) resulted in more precise data and several

papers have noted a lower body fat mass in boys than girls aged 9 to 10 years^{18,19}. Furthermore it was shown, that even in children aged 3 to 8 years girls had significantly more absolute and relative fat mass than boys, while boys exhibited higher amounts of lean soft tissue mass (muscle) and bone mass than girls²⁰. The results of these papers plead for marked gender differences in body composition long before reproductive function makes these differences necessary.

The aim of the present study was to analyze sex differences in body composition and sex differences in the prevalence of overweight in prepubertal children. Furthermore the impact of socioeconomic parameters on growth, weight status and body composition during prepubertal phase of life was tested.

Materials and Methods

Area of research

This research took place in the central part of lower Austria in the south of the river Danube valley. Lower Austria belongs to the eastern Bundesländer of Austria and is a predominantly agricultural region. The administrative center is St. Pölten with 50 026 inhabitants and an altitude of 267m over sea level. Data collection was carried out in 4 primary schools in four different villages (Bischofstetten, Kilb, Hürm, Kirnberg) in the west of St. Pölten in the political district Melk between April and June 1998. The political district Melk is 80km in west the Austrian capital Vienna and a typical agricultural area of the low altitude part of Austria.

Participants

After the approval of the Lower Austrian supervisory school authority in St. Pölten, we contacted school directors and teachers of the primary schools and they sent letters to the parents of the children

to get their approval for the investigation too. 95% of the contacted parents gave their consent to the investigation. Therefore a sample of 398 children (213 girls, 185 boys), ages between 7 and 10, who attended public primary schools in one of the four villages mentioned above was cross sectionally measured. All children were white Europeans with German mother tongue. The sample represented 95% of the 7 to 10 year old children of the four villages under discussion.

Somatometrics and nutritional status

Stature was measured with a Martin anthropometer to the nearest millimeter according to the technique described by Knussmann²¹. Weight was recorded with a scale precise to 100g. The children wore only under wear. Weight status was determined using body mass index (BMI) kg/m^2 . Although the BMI is increasingly used for the diagnosis of obesity, and underweight in childhood and adolescence, up to now there is a lack of European standards. In the present study we used according to the ASNS (Austrian Study of Nutritional Status)²² the 85th percentile as cut-off between normal weight and obesity and the 10th percentile as cut-off between underweight and normal weight for 7, 8 9 and 10 year old children.

Body composition

Body composition was determined using a TBF 105 Body composition analyzer according to BIA-method. In this BIA system two foot-pad electrodes are incorporated into a precision electronic scale. Impedance of the lower extremities and body weight is measured simultaneously while the subject is standing on the scale. The electrodes are in contact with soles and heels of both feet. Biological impedance was measured with 4 terminals and uses a standard 50 kHz – 0.8 mA sine wave constant current. The voltage drop was compared with the heel electrodes. TBF

105 automatically measured weight and impedance. The computer software in the machine then used the measured resistance, the programmed subjects gender, group (child, adult or adult athlete), and stature height and the measured weight to calculate the body density based on previously derived equations obtained from regression analyses with underwater weighing. This was then applied automatically to the standard densiometric formula according to Brozek to calculate the fat percentage. The following variables were determined: Total fat percentage (fat %), absolute Fat mass (in kg) (FM), lean body mass (in kg) (LBM), total body water in kg (TBW). The coefficient of variation for within day impedance measurement was 0.9% and the between days coefficient of variations was 2.1%²³. According to Nunez et al.²³ the leg-to-leg pressure contact electrode BIA system has overall performance characteristics for impedance measurement and body composition analysis similar to conventional arm-to-leg electrode BIA and offers the advantage of increased speed and ease of measurement. The BIA method using bipolar foot electrodes is described as useful and reliable technique for measuring body composition by several studies^{23–26}. In the present study the method was especially useful because the body fat analyzer was easy to transport to the schools where the data collection took place and the children had only went up on the scale and the determination of body weight and body composition were performed simultaneously.

Socioeconomic variables

All children get an anonymous questionnaire which their parents had to fill in, few days before the investigation in the school took place. The questionnaire contained questions regarding educational level and profession of the parents, the age of the parents at child birth and

S. Kirchengast and V. Steiner: Body Composition in Children, Coll. Antropol. **25** (2001) 1: 21–30

native language of the parents. The finished questionnaires were collected together with the somatometric and body composition data.

Statistical analyses

Statistical analyses were carried out using SPSS for Windows 7.0 (Microsoft corp.) After computing descriptive statistics (mean, standard deviation, median) z-scores of the somatometric and body composition variables were computed. Student t-tests were computed to test sexual dimorphism in somatometric and body composition variables. One-way ANOVA (Duncan post hoc tests) were used to assess mean differences in somatometric and body composition z-score by socioeconomic parameters. Cross-tabs (Chi-squares) were computed to test the interaction between sex, age, parental educational level, parental profession and weight status of the children.

Results

Socioeconomic characteristics

The socioeconomic parameters are listed in Table 1. As to be expected in this rural area the majority of father and mothers had finished primary school (until their 14th year of life) only and started with professional training (up to 3 years) immediately after finishing school. Only few parents stayed at school until 18 years and reached high school degree. No parents reached a University degree. Although the research area is a typical agricultural region less than 20% of the fathers worked exclusively as farmers. Nevertheless more than 80% were so called »Nebenerwerbsbauern«. The majority of women (more than 50%) stayed at home as farmer or housewives.

Weight status

As to be seen in Figure 1 and 2 the body mass index of the majority (more

than 70%) of girls and boys corresponded to the definitions of normal weight according to Zarfel and Elmadfa²². Nevertheless significant gender differences in weight status were observed (Chi-square = 4.75, $p < 0.05$). While more than 1% of the boys were classified as underweight only 0.5% of the girls corresponded to the definitions of underweight. In contrast, nearly 30% of the girls were overweight or obese but »only« 20% of the boys corresponded to the definitions of overweight.

Gender differences in growth and body composition

The comparison of stature, body weight and body mass index as indicator for weight status yielded no significant differences between girls and boys. As to be seen in Figure 3 stature which represents skeletal growths shows no gender differences up to 8 years and only small gender

TABLE 1
SOCIOECONOMIC PARAMETERS

Socioeconomic variable	girls	boys
Education of father		
Pflichtschule	84.0%	85.9%
Secondary school	13.5%	9.0%
High school degree	2.5%	5.1%
University degree	0.0%	0.0%
Education of mother		
Pflichtschule	81.9%	79.2%
Secondary school	15.2%	14.0%
High school degree	2.9%	6.7%
University degree	0.0%	0.0%
Profession of father		
Farmer	14.9%	20.3%
Worker	38.8%	39.0%
Employee	36.3%	31.6%
Employer	10.0%	9.0%
Profession of mother		
Farmer (Housewife)	56.4%	57.8%
Worker	10.8%	11.1%
Employee	30.9%	27.2%
Employer	2.0%	3.9%

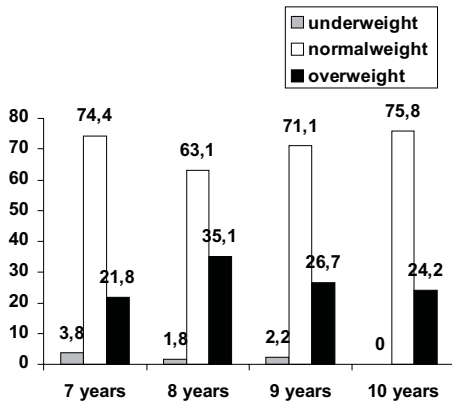


Fig. 1. Weight status in girls by age.

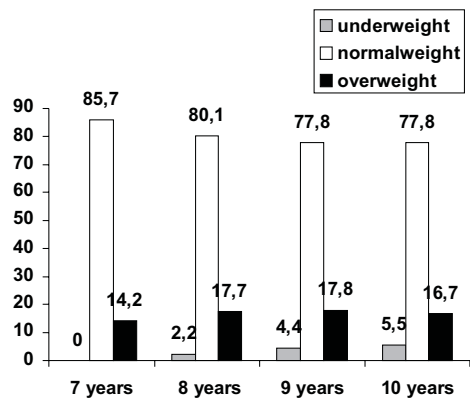


Fig. 2. Weight status in boys by age.

differences between 8 and 10 years. Body weight increased nearly identical in boys and girls between 7 and 10 years. In contrast, the body mass index in 8 to 10 year old girls was higher, however insignificantly higher, than in boys. Although the gender differences in stature, weight and

BMI were insignificant, marked differences occurred in body composition. Girls surpassed boys in absolute and relative fat mass in all age categories, while boys exhibited significant higher values in lean body mass and total body water in each age class (see Table 2, Figures 3–8).

TABLE 2
MEAN VALUES (STANDARD DEVIATIONS) OF STATURE, WEIGHT, ABSOLUTE FAT MASS, FAT PERCENTAGE, LEAN BODY MASS, TOTAL BODY WATER AND BMI FOR GIRLS AND BOYS SEPARATELY BY AGE

Age (years)	N	Stature	Weight	FM (kg)	Fat%	LBM	TBW	BMI
Girls								
7	78	125.0 (5.6)	25.3 (4.6)	5.7** (2.5)	21.7** (5.3)	19.6** (2.4)	14.3** (1.8)	16.2 (2.4)
8	57	132.2 (5.4)	29.7 (5.7)	7.4** (3.2)	23.9** (5.8)	22.4** (2.8)	16.4** (2.1)	16.9 (2.9)
9	45	134.5 (6.1)	31.7 (7.1)	8.4** (4.2)	25.1** (6.8)	23.4** (3.3)	17.1** (2.4)	17.4 (2.9)
10	33	142.2 (6.8)	36.5 (9.1)	9.3* (4.4)	24.4** (5.1)	26.9* (5.1)	19.7* (3.8)	17.9 (3.2)
Boys								
7	56	125.4 (5.4)	26.1 (6.4)	3.9** (2.9)	14.2** (5.9)	22.2** (4.2)	16.2** (3.1)	16.5 (3.0)
8	45	132.2 (7.4)	29.3 (5.8)	4.9** (2.5)	15.8** (5.0)	24.5** (3.6)	17.9** (2.6)	16.6 (1.9)
9	45	136.6 (6.2)	31.3 (6.4)	5.5** (3.9)	16.4** (6.8)	25.9** (3.1)	19.0** (2.3)	16.7 (2.8)
10	36	143.3 (6.4)	36.5 (7.4)	6.8* (3.7)	17.6** (5.9)	29.7* (4.3)	21.7* (3.1)	17.6 (2.6)

FM = fat mass (kg); LBM = lean body mass (kg); TBW = total body water; BMI = body mass index (kg/m²).

* boys' and girls' values differ by < 0.01;

** boys' and girls' values differ by < 0.001.

S. Kirchengast and V. Steiner: Body Composition in Children, *Coll. Antropol.* **25** (2001) 1: 21–30

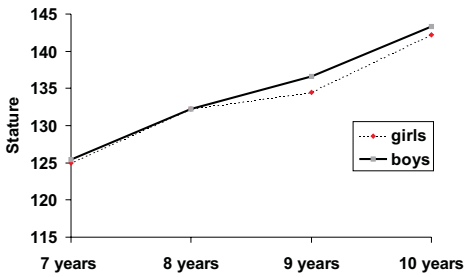


Fig. 3. Stature: cross-sectional.

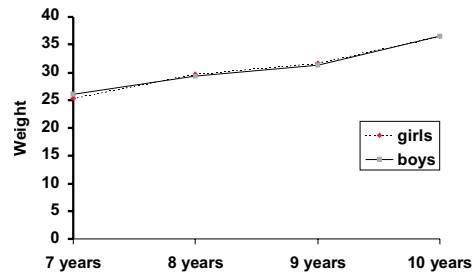


Fig. 4. Weight: cross-sectional.

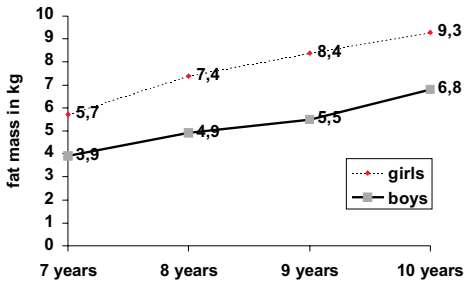


Fig. 5. Fat mass (in kg): cross-sectional.

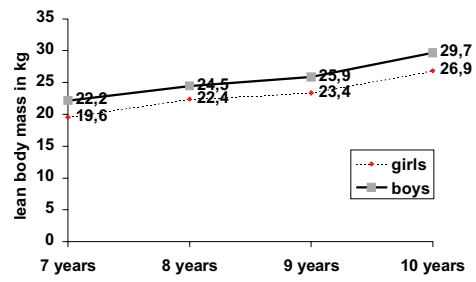


Fig. 6. Lean body mass (in kg): cross-sectional.

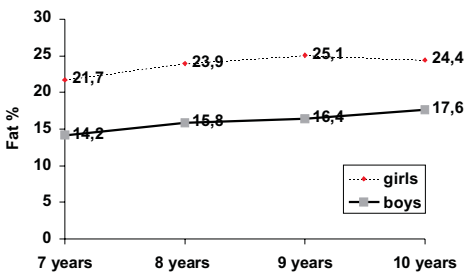


Fig. 7. Fat (%): cross-sectional.

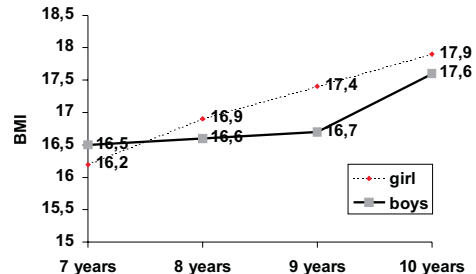


Fig. 8. BMI: cross-sectional.

Socioeconomic factors and growth

Only few significant interactions between socioeconomic variables and skeletal growth were found. With increasing educational level of the mother stature of the offspring increased significantly ($p < 0.05$). No significant associations were found between fathers educational level, fathers or mothers profession and children stature (see Table 3).

Socioeconomic factors and body composition

A statistically significant association between socioeconomic parameters and body composition variables or weight status occurred between lean body mass and mothers education and body mass index and fathers profession only. With increasing educational level of the mother lean body mass increased significantly ($p < 0.03$).

S. Kirchengast and V. Steiner: Body Composition in Children, *Coll. Antropol.* **25** (2001) 1: 21–30

TABLE 3
SOCIOECONOMIC PARAMETERS AND ANTHROPOMETRIC AND BODY COMPOSITION VARIABLES

	N	Stature	Weight	BMI	Fat %	FM	LBM
Father's education							
< High school	320	0.02 (1.01)	0.02 (1.01)	0.03 (1.01)	0.01 (1.00)	0.01 (1.01)	0.02 (1.02)
High school	43	−0.02 (0.88)	−0.15 (0.74)	−0.19 (0.80)	−0.04 (0.92)	−0.12 (0.79)	−0.14 (0.79)
Post secondary	14	−0.15 (1.10)	−0.83 (1.46)	−0.11 (1.40)	−0.33 (1.39)	−0.07 (1.63)	0.07 (1.18)
F-value		0.16 n.s.	0.59 n.s.	0.96 n.s.	0.77 n.s.	0.37 n.s.	0.46 n.s.
Mother's education							
< High school	308	−0.02 (0.99)*	−0.03 (1.00)	−0.03 (0.99)	−0.01 (1.00)	−0.01 (1.01)	−0.03 (0.99)*
High school	56	−0.10 (1.00)*	−0.01 (0.91)	0.13 (1.10)	0.05 (1.11)	0.02 (1.01)	−0.06 (0.92)
Post secondary	18	0.52 (1.14)*	0.43 (1.36)	0.14 (1.09)	−0.09 (0.89)	0.15 (1.17)	0.58 (1.33)*
F-value		2.95 p < 0.05	1.77 n.s.	0.82 n.s.	0.16 n.s.	0.25 n.s.	3.31 p < 0.03
Father's profession							
Farmer	66	−0.09 (0.93)	0.08 (1.05)	0.20 (1.20)*	0.07 (1.08)	0.09 (1.11)	0.05 (1.04)
Worker	147	−0.03 (1.05)	−0.03 (1.06)	−0.06 (0.96)	−0.06 (0.98)	−0.04 (1.01)	−0.02 (1.03)
Employee	129	0.02 (1.01)	−0.08 (0.89)	−0.12 (0.79)*	−0.07 (0.88)	−0.09 (0.81)	−0.05 (0.95)
Employer	36	0.12 (0.92)	0.19 (1.10)	0.19 (1.21)	0.22 (1.21)	0.25 (1.28)	0.12 (1.01)
F-value		0.39 n.s.	0.88 n.s.	2.82 p < 0.05	1.39 n.s.	1.41 n.s.	0.32 n.s.
Mother's profession							
Farmer/ Housewife	219	−0.09 (1.01)	−0.04 (1.08)	0.01 (1.06)	−0.04 (1.01)	−0.02 (1.07)	−0.05 (1.03)
Worker	42	−0.04 (1.09)	−0.06 (0.95)	−0.09 (0.79)	0.01 (0.86)	−0.06 (0.81)	−0.04 (1.06)
Employee	112	0.13 (0.97)	0.04 (0.91)	−0.04 (0.91)	0.01 (0.97)	−0.01 (0.89)	0.07 (0.99)
Employer	11	0.21 (0.59)	0.47 (0.94)	0.63 (1.48)	0.54 (1.52)	0.61 (1.54)	0.31 (0.51)
F-value		1.34 n.s.	1.01 n.s.	1.57 n.s.	1.15 n.s.	1.41 n.s.	0.65 n.s.

Regarding the profession of the father children of farmers showed the highest weight status, while the offspring of employees exhibited the lowest weight status ($p < 0.05$) (see Table 3). No significant associations were found between socioeconomic parameters and absolute and relative fat mass and body weight.

Discussion

Body composition reflects nutritional status and weight status to some extent and is influenced by age, sex, ethnicity, physical activity and disease^{27,28}. From childhood through adolescence to adulthood body composition changes and the degree of this change is sex dependent²⁹. Sexual dimorphism in body composition is evident from adolescence onward throughout adult life and is explained proximate by quantitative gender differences

in sex hormone production³⁰ and ultimate by mechanisms of sexual selection (male-male competition) and natural selection (energetic requirements for successful reproduction). During adolescence boys show a greater increase in lean body mass than girls, who conversely have a greater increase in absolute and relative fat mass^{16,29,31}. The accumulation of additional fat deposits in females during adolescence reflects the enormous energetic demands of reproduction in females³². Fat deposits should be accumulated during puberty or short before and therefore gender differences in body composition should be observable not later than at this stage of development. For a long time prepubertally sexual dimorphisms in body composition was considered to be slight only¹⁶, although at birth there are appreciable sex differences in the thickness of subcutaneous fat tissue and in

S. Kirchengast and V. Steiner: Body Composition in Children, *Coll. Antropol.* **25** (2001) 1: 21–30

overall body mass³³. According to McGowan et al.³³ female babies showed a higher subcutaneous fat mass at five different sites, while boys are on the average 200g heavier than girls at birth. During infancy, childhood and juvenile phase the energetic requirements of boys and girls are more or less equal, they differ not in height and weight. Sex differences in body composition were discussed as relevant not before puberty onset. Nevertheless an increasing number of papers describes gender differences in body composition during early childhood and juvenile phase^{18–20,34–36}. The authors of all these papers described a significant higher amount of body fat in girls than in boys and an increased fat free tissue mass in boys even at ages between 3 and 10 years. The results of the present paper are in accordance with those observations mentioned above. Even long before puberty onset, girls exhibited a higher amount of absolute and relative fat mass, a lower amount of lean body mass although boys and girls did not differ significantly in stature, weight and body mass index. In this way the hypothesis that sex differences in body composition are observable long before the onset of puberty is corroborated by the results of the present study.

Regarding sex differences in weight status categories, we found a significantly higher percentage of overweight or obesity in girls than boys in our sample: Although boys and girls did not differ significantly in stature, weight and body mass index, 29.1% of the girls could be classified as overweight following the weight status definitions of the ASNS²², while this was true of 20% of the boys only. Our results were in accordance with the findings of Cameron and Getz³⁷, who reported an increased prevalence of obesity in African adolescent females and with the findings of Neutzling et al.³⁸ who described a twice as high prevalence of obesity within females than within males

in a Brazilian sample. On the other hand our results were in marked contrast to the findings of Rolland-Cachera et al.³⁹, and Mast et al.³⁵, who described higher body weights and body mass indices in boys. Even the Austrian survey of nutritional status²² yielded a higher prevalence of overweight in body than in girls during childhood and adolescence. Furthermore we have to state another unexpected observation: We found no significant differences in stature between the two sexes. Boys showed only a slightly higher stature than girls even in the age class 9 and 10 years, and this result is rather unexpected: Until about 9 to 10 years boys and girls are very close in height, with boys being very slightly taller and heavier. Growth velocity is nearly identical in girls and boys until the onset of the adolescent growth spurt^{40,41}. The earlier onset of puberty in girls leads to an acceleration of height at 9 or 10 years and so to higher statures in girls at this age. Our results with a slightly increased stature even in 10 year old boys is therefore in contrast to the descriptions of Prece⁴⁰ and Bogin⁴¹. The slightly higher stature and the significantly lower amount of body fat observed in boys is also in contrast to the findings of Himes and Roche⁴², who reported that obese children are of above average stature and that stature is positively correlated with the amount of subcutaneous fat tissue during prepuberty. These unexpected results, which are not in accordance with previous studies, especially with the results of the Austrian survey of nutritional status²² demonstrates the limitations of possible interpretations of the results of this study. We have to be aware that we only analyzed a relatively small sample from a small rural area of Austria and the results are therefore not representative for Austria in general.

Acknowledgements

We are gratefully indebted to the Lower Austria supervisory school authority and the directors and teachers of the primary school in Kilb, Bischoffstetten, Hürm

and Kirnberg for their approval to the present study. Special thanks go to the school children and their parents without their cooperation our study could not take place.

REFERENCES

1. LIVINGSTONE, B., *Eur. J. Pediat.*, 159 (2000) S24. — 2. BJÖRNTORP, P., *Lancet*, 350 (1997) 423. — 3. BRIEFEL, R., M. McDOWELL, K. ALAIMO, C. CAUGHMAN, A. BISCHOF, M. CARROLL, C. L. JOHNSON, *Am. J. Clin. Nutr.*, 62 (1995) 1072S. — 4. FREEDMAN, D. S., S. R. SRINIVASAN, R. A. VALDEZ, D. F. WILLIAMSON, G. S. BERENSON, *Pediatrics*, 99 (1997) 420. — 5. FLEGAL, K. M., R. P. TROIANO, *Int. J. Obes. Rel. Metab. Dis.*, 27 (2000) 807. — 6. ARMSTRONG, N., J. BALDING, P. GENTLE, B. KIRBY, *Brit. Med. J.*, 301 (1990) 203. — 7. BUCHOWSKI, M. S., M. SUN, *Int. J. Obes.*, 20 (1996) 236. — 8. KATZMARZYK, P. T., R. M. MALINA, T. M. SONG, C. BOUCHARD, *J. Adolesc. Health*, 23 (1998) 318. — 9. KATZMARZYK, P. T., R. M. MALINA, C. BOUCHARD, *Prev. Med.*, 29 (1999) 555. — 10. ROLLAND-CACHERA, M. F., M. DEHEEGER, M. AKROUT, F. BELLISLE, *Int. J. Obes. Rel. Metab. Dis.*, 19 (1995) 573. — 11. SERDULA, M. K., D. IVERRY, R. J. COATES, D. S. FREEDMAN, D. F. WILLIAMSON, T. BYERS, *Prevent. Med.*, 22 (1993) 167. — 12. DIETZ, W. H., 1998, *J. Nutr.*, 128 (1998) 411S. — 13. ROLLAND-CACHERA, M. F., F. BELLISLE, M. DEHEEGER, F. PEQUIGNOT, M. SEMPE, *Int. J. Obes.*, 14 (1990) 473. — 14. MALINA, R. M., C. BOUCHARD: *Subcutaneous fat distribution during growth*. (Alan Liss, New York, 1988). — 15. MALINA, R. M., S. KOZIEL, T. BIELICKI, *Am. J. Hum. Biol.*, 11 (1999) 189. — 16. FORBES, G.: *Human body composition: Growth, ageing, nutrition and activity*. (Springer Verlag, New York, 1987). — 17. KELLY, J., W. STANTON, R. MCGEE, P. SILVA, *J. Ped. Child Health*, 28 (1992) 158. — 18. FAULKNER, R. A., D. A. BAILEY, D. T. DRINKWATER, A. A. WILKINSON, C. S. HOUSTON, H. A. MCKAY, *Calcif. Tissue Int.*, 53 (1993) 7. — 19. NELSON, D., P. SIMPSON, C. JOHNSON, D. BARONDESS, M. KLEEREKOPER, *Bone*, 20 (1997) 73. — 20. TAYLOR, R. W., E. GOLD, P. MANNING, A. GOULDING, *Int. J. Obes.*, 21 (1997) 1082. — 21. KNUSSMANN, R.: *Somatometrie*. (Gustav Fischer Verlag, Stuttgart, 1988). — 22. ZARFL, B., I. ELMADFA, *Akt. Ernährungsmed.*, 20 (1995) 201. — 23. NUNEZ, C., D. GALLAGHER, M. VISER, F. X. PI-SUNYER, X. WANG, S. B. HEYMSFIELD, *Med. Sci. Sports Ex.*, 29 (1997) 524. — 24. TSUI, E. Y. L., X. J. GAO, B. ZINMAN, *Diab. Med.*, 15 (1998) 125. — 25. HAINER, V., M. KUNESOVA, J. PARZIKOVA, V. STICH, J. HOREJS, L. MULLER, *Sb. Lek.*, 96 (1995) 249. — 26. XIE X., N. KOLTHOFF, O. BARENHOLT, S. P. NIELSEN, *Int. J. Obes. Relat. Dis.*, 23 (1999) 1079. — 27. ROLLAND-CACHERA, M. F., *Horm. Res.*, 39 (1993) 25. — 28. KATZMARZYK, P. T., M. C. MAHANEY, J. BLANGERO, J. QUEK, R. M. MALINA, *Hum. Biol.*, 71 (1999) 977. — 29. CHUMLEA, W. C., R. M. SIERVOGEL, A. F. ROCHE, P. WEBB, E. ROGERS, *Hum. Biol.*, 55 (1983) 845. — 30. SHEPARD, R. J.: *Body composition in biological anthropology*. (Cambridge Studies in Biological Anthropology, Cambridge University Press, Cambridge, 1991). — 31. CHUMLEA, W. C., J. L. KNITTLE, A. F. ROCHE, R. M. SIERVOGEL, P. WEBB, *Am. J. Clin. Nutr.*, 34 (1981) 1791. — 32. FRISCH, R. E., *Perspect. Biol. Med.*, 28 (1985) 611. — 33. MCGOWAN, A., M. JORDAN, J. MACGREGOR, *Biol. Neonatol.*, 25 (1975) 66. — 34. BOOT, A. M., J. BOUQUET, M. A. DE RIDDER, E. P. KRENNING, S. M. DE MUINCK KEIZER-SCHRAMA, *Am. J. Clin. Nutr.*, 66 (1997) 232. — 35. MAST, M., I. KORTZINGER, E. KONIG, M. J. MULLER, *Int. J. Obes.*, 22 (1998) 878. — 36. MOLLGAARD, C., K. F. MICHAELSEN, *Appl. Rad. Isotop.*, 49 (1998) 577. — 37. CAMERON, N., B. GETZ, *Int. J. Obes. Relat. Metab. Dis.*, 21 (1997) 775. — 38. NEUTZLING, M. B., J. A. TADDAI, E. M. RODRIGUES, D. M. SIGULEM, *Int. J. Obes. Rel. Metab. Dis.*, 24 (2000) 869. — 39. ROLLAND-CACHERA, M. F., T. J. COLE, M. SEMPE, J. TICHET, C. ROSSIGNOL, A. CHARRAUD, *Eur. J. Clin. Nutr.*, 45 (1990) 13. — 40. PREECE, M. A., *Sexual dimorphism of physique*. In: ULJASZEK S. J., F. E. JOHNSTON, M. A. PREECE (Eds.): *The Cambridge encyclopedia of human growth and development*. (Cambridge University Press, Cambridge, 1998). — 41. BOGIN, B.: *Patterns of human growth*. (Cambridge University Press, Cambridge, 1999). — 42. HIMES, J. H., A. F. ROCHE, *Hum. Biol.*, 58 (1986) 737.

S. Kirchengast and V. Steiner: Body Composition in Children, *Coll. Antropol.* **25** (2001) 1: 21–30

S. Kirchengast

Institute for Anthropology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria

SPOLNI DIMORFIZAM U GRAĐI TIJELA, TJELESNOJ MASI I VISINI U ŠKOLSKE DJECE PRETPUBERTETSKE DOBI IZ RURALNIH PODRUČJA ISTOČNE AUSTRIJE

S A Ž E T A K

Spolni dimorfizam u visini tijela, tjelesnoj masi te građi tijela analiziran je na uzorku od 398 pretpubertetske djece (213 djevojčica, 185 dječaka) stare od 7 do 10 godina. Prevalencija pretilosti među djecom ovog uzrasta također je testirana. Parametri građe tijela određeni su korištenjem TBF analizatora tjelesne građe prema BIA metodi. Opažene su visoko značajne razlike u tjelesnoj građi ($p < 0.001$) između dječaka i djevojčica. Nasuprot tome, visina tijela, tjelesna masa i BMI nisu pokazali značajne razlike među spolovima. Značajno ($p < 0.05$) veći postotak djevojčica (29%) u odnosu na dječake (20%) pripada kategoriji pretilih, prema definiciji pretilosti ASNS-e (Austrian Survey of Nutritional Status). Rezultati ove studije pokazali su ne samo značajnu spolnu razliku u građi tijela (posebice u količini masnog tkiva) znatno prije početka puberteta, već i značajno višu prevalenciju pretilosti u pretpubertetskih djevojčica u odnosu na dječake.



ON THE EVOLUTION OF THE SEX DIFFERENCES IN THROWING:
THROWING IS A MALE ADAPTATION IN HUMANS

MICHAEL P. LOMBARDO

*Biology Department, Grand Valley State University
Allendale, Michigan 49401-9403 USA*

E-MAIL: LOMBARDM@GVSU.EDU

ROBERT O. DEANER

*Psychology Department, Grand Valley State University
Allendale, Michigan 49401-9403 USA*

E-MAIL: ROBERT.DEANER@GMAIL.COM

KEYWORDS

anatomy, combat, fighting, human evolution, hunting, throwing, sex differences

ABSTRACT

The development of the ability to throw projectiles for distance, speed, and accuracy was a watershed event in human evolution. We hypothesize that throwing first arose in threat displays and during fighting and later was incorporated into hunting by members of the Homo lineage because nonhuman primates often throw projectiles during agonistic interactions and only rarely in attempts to subdue prey. Males, who threw more often than females in both combat and hunting, would have been under stronger selection than females to become proficient at the ability to throw, intercept, and dodge projectiles as throwing skills became critical to success in combat and hunting. Therefore, we predict that males, more than females, should display innate anatomical and behavioral traits associated with throwing. We use data from a variety of disciplines to discuss: the sex differences in throwing speed, distance, and accuracy; sex differences in the development of the throwing motion; inability of training or cultural influences to erase the sex differences in throwing; sex differences in the use of throwing in sports, combat, and hunting; and sex differences in anatomical traits associated with throwing that are partly responsible for male throwing superiority. These data contradict the view held by many commentators that socialization rather than innate sex differences in ability are primarily responsible for male throwing superiority. We suggest that throwing is a male adaptation.

INTRODUCTION

“THROWING LIKE A GIRL”

GIRLS and women typically do not throw as fast, far, or accurately as boys and

men (Thomas and French 1985). The reasons for these differences have been the subject of debate for decades. Many scholars and commentators in the popular media have claimed that the sex differences in throwing are not as

The Quarterly Review of Biology, June 2018, Vol. 93, No. 2

Copyright © 2018 by The University of Chicago Press. All rights reserved.

0033-5770/2018/9302-0002\$15.00

large as the public believes and that they are the result of sex differences in socialization and not differences in innate abilities (Table 1). Indeed, the commentary on this topic often includes statements such as: the pejorative trope to “throw like a girl” has at its “source in the particular *situation* of women as conditioned by their sexist oppression in contemporary society. . . . Women in sexist society are physically handicapped” (Young 1980:152); “‘throwing like a girl’ is quite likely to be a recent, idiosyncratic, culturally imposed sexual dimorphism” (Bingham and Souza 2010); and “Young women learning to throw in North America thus face a quadruple whammy of obstacles. We *are* girls, therefore we throw like ourselves. We are learners, therefore our bodies naturally perform the early throwing mechanics that have been culturally labeled in a negative way. Even post-*Title IX*, many of us come to this experience later and have fewer opportunities to practice overhand throwing. So our early-learning phase lasts longer and looks more acute by contrast with age-cohort boys, who have moved through those early learner phases more quickly. All this is a powerful disincentive to keep picking up the ball. Because throwing is such a complex, learned activity, . . . this vicious cycle can have a disproportionately large effect” (Rowe 2012).

Few commentators have discussed at length the possibility that the origin of the sex differences are innate differences between males and females that have been shaped by selection. The view of many feminists, journalists, political theorists, and scientists, including evolutionarily oriented ones, is that girls “throw like girls” *only* because of sex differences in socialization (Table 1). However, from an evolutionary perspective this view is fatally flawed because it explicitly rejects our modern understanding about how an individual’s traits are expressed. This is because throwing is a phenotype and like all other phenotypes arises from the complex interactions between an individual’s genotype and its environment (Plomin et al. 2008). Therefore, it is surprising that few scholars and commentators have hypothesized that the sex differences in throwing have partly innate causes (Table 1).

WHY THERE ARE SEX DIFFERENCES IN THROWING IS AN INTERESTING QUESTION

It is not surprising that males throw faster and farther than females for two related reasons. First, adult males are larger than females (Ellis et al. 2008). The length of the body and limbs synergistically contribute to throwing velocity (Stodden et al. 2006a,b; Maki 2013; Roach and Lieberman 2014). Second, males have more muscle mass in both absolute terms and relative to fat-free body mass (Janssen et al. 2000; Lassek and Gaulin 2009) and consequently are stronger than females. Muscle mass contributes to the amount of force that can be used during the throwing motion. Numerous studies comparing athletically trained and untrained males and females have shown that female strength, even when controlling for differences in fat-free body mass and limb or body size, ranges from 40–75% of male strength (Bishop et al. 1987; Lassek and Gaulin 2009). Especially relevant to the sex difference in throwing velocity, the difference is greater for upper-body than lower-body strength (Bishop et al. 1987; Janssen et al. 2000; Lassek and Gaulin 2009). Moreover, male muscle fibers are shorter and have greater angles of pennation (i.e., the angle of the muscle fibers to the force-generating axis of the muscle) contributing to the sex differences in muscular strength (Chow et al. 2000). Finally, males tend to have a greater proportion of “fast twitch” (Type IIA) muscle fibers than females (Miller et al. 1993) making male muscles better at producing the anaerobic power (Mayhew and Salm 1990) used during the rapid, intense actions that occur during throwing. Thus, the larger size and greater strength of males are responsible for some of the sex differences in throwing. However, what makes the sex differences in throwing especially interesting is that they are greater than the sex differences in running and jumping.

A comparison between male and female performances in track and field events demonstrates the large gap in male and female overhand throwing performance relative to the other basic physical actions of running and jumping. The sex differences in world records, top 10 all-time performances, the per-

TABLE 1
A sample of some published discussions about the evolution of throwing

No discussion of sex differences in throwing	Sex differences in throwing attributed solely to environmental effects	Sex differences in throwing attributed to biology x environment interaction
Darlington (1975)	Young (1980, 1998)	Thomas and French (1985)
Calvin (1983)	Fallows (1996)	Isaac (1987)
Knüsel (1992)	Williams (1996)	Thomas and Marzke (1992)
Young (2003)	Dunsworth et al. (2003)	Ehl et al. (2005)
Bingham and Souza (2009)	Evaldsson (2003)	Young (2009)
Roach et al. (2012, 2013)	Fredrickson and Harrison (2005)	Downey (2010)
Roach and Lieberman (2014)	Anthamatten (2014)	
Roach and Richmond (2015)		
Larson (2015)		

formances of 2016 Olympic finalists, and U.S. youth record holders show that males, on average, run about 10% faster and jump about 15–20% higher than females (Table 2). There are four throwing events: shot put, discus, hammer, and javelin. Males throw heavier implements than females in each event (Table 2). The sex difference in javelin throwing is the most relevant because the javelin is the only implement thrown overhand. In general, there is little difference between males and females in the distances that the shot put, discus, and hammer are thrown, but boys and men throw the javelin more than 30% farther (Table 2). Given that the difference between the weights of the javelins thrown by the sexes is relatively small compared to the differences in weights between the shot put, hammer, and discus thrown by the sexes (Table 2), we would expect there to be relatively little difference in the distance that the sexes throw the javelin if male and female throwing abilities were roughly equivalent. The large sex difference in javelin throwing is *not* because male champion javelin throwers are from the U.S. where overhand throwing is highly valued because of the cultural importance of baseball and softball to Americans. The senior male world record holder is a European as are eight of the top 10 all-time performers (the top eight are Europeans, a Kenyan is ranked ninth, an American is ranked 10th), and none of the top eight finishers in the 2016 Olympics were Americans (six of eight were Europe-

ans, the two others were a Kenyan and a competitor representing Trinidad and Tobago; <https://www.iaaf.org>; <https://www.olympic.org/rio-2016/athletics/javelin-throw-men>). Interestingly, the sex difference for throwing baseballs for speed is similar to that for javelin throwing for distance; men throw baseballs 39% faster (169 km/hr versus 105 km/hr) than women (Glenday 2007). Because they are throwing the same weight implement (range 141.75–148.83g), the difference in baseball throwing speed clearly illustrates the sex difference in throwing performance. In summary, the sex difference in javelin throwing performance is larger than those in running and jumping and the magnitude of the difference is paralleled by the sex difference in throwing baseballs for speed. Collectively, these patterns suggest that there may be a fundamental difference between the sexes in throwing ability that goes beyond the differences in size and strength.

AN EVOLUTIONARY THEORY OF THROWING

Our overarching purpose is to develop an evolutionary theory of throwing in humans. Elsewhere, we hypothesized that throwing likely evolved first within the ecological context of agonistic interactions and then transitioned into use during hunting because throwing in nonhuman primates occurs almost exclusively during threat displays and agonistic interactions (Lombardo and Deaner

TABLE 2
Sex differences in performance in track and field events

Event	Percent sex difference in senior world record performances	Percent sex difference in senior all-time top 10 performances	Percent sex difference in 2016 Olympic final performances	Percent sex difference in U.S. youth performances
Running events				
100m	8.7	8.8	9.4	6.5
200m	10.1	9.8	9.5	6.2
400m	9.6	10.2	12.6	8.0
800m	10.9	11.1	11.5	7.7
1500m	10.5	10.6	8.3	7.6
3000m	9.3	10.3	— ³	6.6
5000m	11.0	11.1	11.1	14.6
10,000m	10.2	11.0	9.3	12.7
Marathon	10.2	11.0	9.9	— ⁴
Jumping events				
High jump	17.2	17.0	20.1	12.5
Pole vault	21.3	23.4	21.4	23.7
Long jump	19.0	18.6	18.6	13.8
Triple jump	18.0	17.8	16.6	18.3
Throwing events ¹				
Shot put	2.2	3.4	9.5	23.3
Discus	3.5 ²	2.4 ²	1.5	25.0
Hammer throw	4.5	8.4	2.6	18.6
Javelin	36.2	33.2	31.9	33.7

All data from International Association of Athletics Federations (<https://www.iaaf.org>) or USA Track & Field (<http://usatf.org>). Data represented are the percent differences between male and female performances; instances where female performance is superior to male performance are indicated in the table. Sex differences in top 10 all-time performances based on mean performance in each event, 2016 Olympic final based on mean performance by the eight finalists in each event, and U.S. youth athletic performances are based on the mean sex differences in each event averaged across all age groups; male performances were superior to female performances in all events across all youth age groups. U.S. youth age groups are <= 8, 9–10, 11–12, 13–14, 15–16, and 17–18 years. The percent sex difference in the javelin throw is in bold font to emphasize the magnitude of the difference compared to the differences in male and female performances in other events.

¹Implements used by males are heavier than those used by females. Senior throwing implements: male shot = 7.26kg; female shot = 4kg; male discus = 2kg; female discus = 1kg; male hammer = 7.26kg; female hammer = 4kg; male javelin = 800g and 2.6–2.7m in length; female javelin = 600g and 2.2–2.3m in length. As youth age, the implements thrown by girls become lighter than those thrown by boys. Shot: ages 8 and under = 2kg; ages 9–12 = 2.72kg; ages 13–14, boys = 3.63kg, girls = 2.72kg; ages 15–18, boys = 5.44kg, girls = 4kg. Discus: ages up to 10 depends on local rules; ages 11–14 = 1kg; ages 15–18, boys = 1.6kg, girls = 1kg. Javelin: ages 8–12 = 300g; ages 13–14 = 600g; ages 15–18, boys = 800g, girls = 600g (<http://www.usatf.org/groups/officials/files/resources/weights-and-measures/Throws-Implement-Specs-Feb2014.pdf>).

²Female performance is superior to male performance.

³3000m event is not contested in Olympics.

⁴Not an official U.S. youth event.

2018). Here we focus on the evolutionary consequences of over two million years of selection on the biology of throwing in the human lineage (Roach et al. 2013). In this article, we demonstrate the weakness of the claims that socialization alone is responsible for the sex differences in throwing by showing that

the differences are large and primarily due to the evolution of male-specific anatomical and behavioral adaptations that evolved during a long history of stronger selection on the throwing biology of males than on females. First, we review our hypothesis that the ecological causes of selection that shaped the evo-

lution of throwing were stronger on males than on females. This hypothesis yields the predictions that males should be superior throwers than females and that males, more than females, should possess innate behavioral and anatomical traits associated with throwing. Second, we briefly review the evidence showing that males are superior throwers and the developmental and cross-cultural literature reporting sex differences in throwing behavior. Finally, we review the anatomical literature that demonstrates that there are sex differences in the anatomical traits associated with throwing and shows that some of these differences cannot be attributed solely to sex differences in usage.

Others have documented some of the data that we discuss here, including those that show that there is a substantial sex difference in throwing performance, the sex difference is relatively impervious to environmental input, and anatomical traits advantageous for throwing are more pronounced in males than females. However, they did not use these data to develop a complete evolutionary theory of throwing. In contrast, we use several lines of converging evidence (i.e., a consilience of inductions; Wilson 1998) from a variety of disciplines to make a strong argument that the traits underlying the sex difference in throwing are evolved consequences of a long history of selection on male throwing ability.

THE ECOLOGICAL CAUSES OF SELECTION THAT SHAPED THE EVOLUTION OF THROWING ABILITY WERE STRONGER ON MALES THAN ON FEMALES

Several lines of evidence suggest that the ecological causes of selection shaping the evolution of throwing were stronger on males than on females. First, throwing during agonistic interactions is widespread in nonhuman primates (Shumaker et al. 2011). More importantly, male chimpanzees throw more often than females during agonistic interactions (Sugiyama and Koman 1979; Goodall 1986). For example, male chimpanzees at Gombe, Tanzania, were responsible for 94% of throwing incidents recorded in 1968 and 1977–1982 (Goodall 1986). These observations are especially relevant to our theory because it sug-

gests that male hominins were more likely than female hominins to use throwing to manipulate the behavior of other individuals. Thus, selection would have more strongly favored the evolution of superior throwing skills in male, rather than female, hominins if those skills led to increased success during fighting or dominance displays that increased their access to food or mates.

Second, throwing and projectile use during primitive (i.e., without firearms) and modern warfare are common (van Creveld 1989; Isaac and Isaac 2011). Indeed, one of the most common sex differences is that males more often than females throw projectiles during warfare (Murdock and Provost 1973; Wood and Eagly 2002). Throughout human history, lethal violence has been common (Gómez et al. 2016) and males have been far more likely than females to fight and participate in warfare (Turney-High 1971; Wilson and Daly 1985; Keeley 1996; Puts 2010). These circumstances suggest that there would have been stronger selection for throwing abilities in males than in females. Moreover, the effects of selection on throwing ability generated by frequent warfare were likely very strong because of the effects warfare can have on male survival. Archeological and ethnographic data suggest that 15–30% of male mortality in some groups may be caused by warfare and its consequences (Keeley 1996; Otterbein 2004). These levels of male mortality during primitive warfare suggest intense selection on the physical attributes, including throwing ability, that would promote success during combat and fighting (Lombardo and Deaner 2016). The advantages of being skilled at throwing, the use of projectile weaponry, and avoiding projectiles during combat are obvious and would lead to survival and reproductive advantages. Successful fighters and warriors are often rewarded with material goods and achieve high status, and are thereby able to gain greater access to fertile women, and thus obtain greater reproductive success, than nonwarriors (Chagnon 1988; Zerjal et al. 2003) resulting in selection for throwing ability and projectile use.

Finally, throwing projectiles is common in primitive hunting, especially for large game (Oswalt 1976; Shea 2009). Because hunting

large game with thrown projectile weapons has been and remains primarily a male activity (Murdock and Provost 1973; Marlowe 2007), selection on throwing abilities would have been stronger on males than on females. Men who are good hunters, and therefore likely to be highly skilled at using projectiles, typically obtain high status and tend to have high reproductive success (Smith 2004; Gurven and Hill 2009) further augmenting selection for throwing ability.

Collectively, the adaptive advantages of being successful during fighting and as a hunter would have accelerated the evolution of superior male skills at throwing and projectile use because of the synergistic interaction between natural selection driven by the outcomes of male-male competition during combat and hunting and sexual selection driven by female mate choice for superior warriors and hunters. Moreover, the evolution of superior throwing abilities would have been rapid if there were large differences of fitness between individuals that had heritable differences in throwing ability (Fisher 1958).

EVIDENCE OF SEX DIFFERENCES IN THROWING ABILITY

The sex differences in throwing include differences in throwing velocity, distance, and targeting ability (Kimura 1999) and are among the largest documented behavioral sex differences (Thomas and French 1985; Hyde 2005). The sex differences in adult throwing ability are not strongly influenced by sex differences in the amounts of sports activity, especially throwing games, undertaken (Watson and Kimura 1991) and extensive practice does not eliminate them (Duffy et al. 2004).

MALES THROW FASTER THAN FEMALES

At all ages, males throw faster, on average, than females. The sex differences in throwing velocity emerge during the preschool years before sex differences in sports involvement appear. By ages 4–7, 90% of boys throw faster than the average of same-age girls (Thomas and French 1985) and boys 6 years of age

throw faster, on average, than do girls who are 9 years old (Rippe et al. 1990). Practice improves female throwing velocity but, when compared to males with similar throwing experience, the sex difference remains (Thomas and Marzke 1992; Petranek and Barton 2011). By age 12, the fastest throwing girls are comparable to the slowest throwing boys (Thomas and French 1985). Sex differences in throwing velocity persist through adolescence into middle age and although the differences decrease in the oldest age groups, they remain large (Lorson et al. 2013).

MALES THROW FARTHER THAN FEMALES

Throwing distance is positively and highly correlated with throwing speed and is influenced by the point in the throw when the projectile is released (Axe et al. 1996, 2014; Zhu et al. 2009). The magnitude of the sex differences in the development of distance throwing exceeds that of any other motor skill (Espenschade 1960) and manifests itself early; boys 6 years of age throw farther than girls who are 9 years old (Rippe et al. 1990), and only the very best girls throw as far as the least skilled boys at age 17 (Thomas and French 1985). Practice improves throwing distance for both sexes (Zhu et al. 2009), but females do not “catch up” to males (Thomas et al. 1994). The inability of girls to catch up to boys suggests that there may be a biological component that limits the ability of females to improve their throwing with training (Hauenstricker and Seefeldt 1986).

The sex differences in throwing the javelin demonstrate that highly trained women, even when they throw lighter projectiles, do not throw as far as men (Table 2). It is interesting to note that the javelin is most weapon-like of all of the projectiles thrown during track and field competitions.

The difference between the sexes in throwing for distance has consequences during warfare and hunting. Because males can throw farther than females, male warriors and hunters need not approach rival combatants or dangerous prey, respectively, as closely as females would thereby reducing their risks of detection and injury. The potential conse-

quences of sex differences in throwing ability remain relevant in modern warfare (Browne 2007). For example, men in the U.S. Army are required to throw fragmentation grenades 35m whereas women are required to throw them 25m; the blast radius of the grenades is 15m (Department of the Army 2005). Therefore, a toss of only 60% of the required distance could have lethal consequences for female but not male grenade throwers.

MALES HAVE BETTER AIMED THROWING THAN FEMALES

The ability to throw a projectile and hit a target would have been an important cause of selection on ancient warriors and hunters. The sex difference in targeting ability while throwing manifests itself early and persists throughout life (Jardine and Martin 1983; Watson and Kimura 1991). Overall, 75% of men outperform the average woman at hitting a target with a thrown object (Watson and Kimura 1991). Differential sports experience with unimanual games, especially throwing games, contributes only a small proportion to the sex difference in throwing accuracy (Watson and Kimura 1988, 1991). This difference in throwing accuracy can have sex-biased fitness consequences in warfare and hunting because it makes males more effective than females as warriors and hunters.

MALES ARE BETTER AT INTERCEPTING PROJECTILES THAN FEMALES

The advantages of being able to intercept or dodge projectiles during combat are obvious. Males are better, on average, than females at judging the velocity of moving objects (Schiff and Oldak 1990; Law et al. 1993; Sanders and Sinclair 2011), which is advantageous in judging the expected location of a projectile targeting them. Males outperform females at hitting a moving target with a thrown ball, although physical strength affecting accuracy at hitting distant targets and experience may be responsible for some of the sex difference (Peters 1997). The intercepting abilities of both males and fe-

males improve with practice and feedback, but the sex difference remains (Law et al. 1993).

SEX DIFFERENCES IN THE DEVELOPMENT OF THROWING ABILITY

Throwing begins early in life when infants begin to manipulate objects. The first attempts at throwing are usually underhand tosses that eventually transition to overhand throwing (Cratty 1979). Boys tend to achieve mature throwing actions before girls (Butterfield and Loovis 1993; Marques-Bruna and Grimshaw 1997). Throwing proficiency involves flexion and extension of the trunk and arm with weight shifting from the foot on the same side as the throwing arm to weight shifting to the foot opposite the throwing side and with lower trunk and pelvis rotation occurring before upper trunk rotation (Williams and Monsma 2007). For example, boys, on average, show throwing proficiency at 69 months of age while girls do not show throwing proficiency until nearly three years later at 102 months of age (Williams and Monsma 2007). By age 3, boys throw much faster, farther, and more accurately than girls (Thomas and French 1985). There is no other motor skill early in life, middle childhood, or adolescence that boys perform so much better than girls. Before adolescence there is little sex difference in height, weight, and muscle mass (National Center for Health Statistics 2017), so male superiority in throwing cannot be attributed to sex differences in size or strength.

Many scholars of throwing attribute this developmental sex difference in throwing to greater male participation in ball sports and social encouragement of throwing skills by boys (Cratty 1979; Sakurai and Miyashita 1983; Bingham and Souza 2009). However, Butterfield and Loovis (1993) showed that sex differences in the throwing motion and performance remained after age and Tee-ball (simplified baseball), softball, or baseball participation were statistically controlled during analyses. These results are consistent with our hypothesis that the sex difference in throwing may have a biological component.

Our hypothesis predicts sex differences in throwing by children even in cultures where overhand throwing is not encouraged. Cross-culturally, many sex differences in interests and activities emerge before 2 years of age and continue to manifest themselves throughout childhood, adolescence, and adulthood (Cohen-Bendahan et al. 2005; Berenbaum 2011; Ceci and Williams 2011) and are related to adult reproductive strategies (Low 1989, 2000; Geary 1999). Some of these differences, including observations that boys show greater interest than girls in imitating propulsive actions, playing with weapons, and playing hunting and war games (Hoffman 1890; Eibl-Eibesfeldt 1989; Goldstein 1995; Loy and Hesketh 1995; Chagnon 1997; Benenson et al. 2011; Deaner and Smith 2013) likely reinforce the innate sex differences in throwing.

ENVIRONMENTAL INFLUENCES HAVE LITTLE EFFECT ON SEX DIFFERENCES IN THROWING

The overhand throwing motion is a human universal; it develops naturally and appears before children learn to walk (Wickstrom 1977). The early onset of throwing, the pattern of development of the throwing motion, the age-related improvement in throwing ability, and the development of sex differences in throwing ability have been recorded wherever they have been studied, including Brazil, England, Germany, Japan, Mexico, New Guinea, Nigeria, Senegal, Tasmania, and the U.S. (Young 2009).

Male infants may have a greater predisposition for throwing than female infants because they imitate more frequently the propulsive movements (e.g., pushing, throwing, and kicking) associated with throwing (Benenson et al. 2011). Benenson et al. (2011) speculated that these forceful propulsive actions are behavioral expressions of masculinity and that masculine children (i.e., mostly boys) may derive greater pleasure from them than do nonmasculine children (i.e., mostly girls). Research in nonhuman primates and prenatally androgenized girls, (e.g., girls with Congenital Adrenal Hyperplasia) support the hypothesis that prenatal androgen exposure

predisposes males to performing propulsive actions such as throwing. These observations led Benenson et al. (2011) to argue that this innate predisposition, and the pleasure that males derive from it, may provide the foundation for the male preference for games involving projectiles.

TRAINING DOES NOT ELIMINATE THE SEX DIFFERENCES IN THROWING MECHANICS

There is individual variation in the effects of training on the development of throwing skills (Langendorfer and Roberton 2002). Individuals who reach the most advanced developmental levels for the major components of expert and efficient throwing are almost always boys (Stodden et al. 2006a, b). Boys tend to more quickly develop the mature throwing motion because they play more ball games than girls (McKenzie et al. 2002; Vandaele et al. 2011). However, girls and women, even with training, do not often display the most advanced components of the mature throwing motion (Stodden et al. 2006a, b), although training can produce improved throwing motions in preschool girls (Lorson et al. 2013; Veldman et al. 2017). These observations suggest the possibility that innate factors may limit the ability of girls to produce the mature throwing motion.

CULTURE HAS LITTLE EFFECT ON SEX DIFFERENCES IN THROWING

Cultural norms regarding sports, hunting, and warfare have little effect on sex differences in throwing. Thomas et al. (2010) studied throwing in Australian Aboriginal boys and girls aged 6, 8, and 10 years. Both male and female Australian Aborigines throw during hunting and defense (Clarke 2003) leading Thomas et al. (2010) to predict that there would be few sex differences in throwing in this population. However, contrary to their prediction, boys still outperformed girls and the mean effect sizes were large ($d = 1.6$; Cohen 1992) but smaller than in same-aged populations in the U.S. (mean effect size, $d = 2.3$; Thomas et al. 2010). Australian Aborigi-

nal girls outperformed girls from Germany and the U.S. indicating that throwing experience was responsible for their superior performance (Thomas et al. 2010). Indeed, Australian Aboriginal girls threw balls at 78% of the velocity of those thrown by Australian Aboriginal boys, the smallest sex difference in ball velocity recorded in non-Aboriginal Australian, German, Japanese, Thai, and U.S. populations (Young 2009; Thomas et al. 2010). The throwing abilities of the Aboriginal boys were comparable to boys from elsewhere.

Sex differences in throwing also occur in cultures where throwing games are not popular. Ehl et al. (2005) compared the throwing velocity of 13-year-old boys and girls in the U.S. and Germany. Throwing sports and games are more popular with boys and girls in the U.S. than they are in Germany where the most popular sport that involves frequent overhand throwing is team handball, which is not very popular compared to soccer. Ehl et al. (2005) predicted that sex differences in throwing would be smaller in Germany than in the U.S. where there are more sex differences in throwing experience. They found sex differences in both populations and the relative differences between the sexes were about the same. In terms of effect sizes, and contrary to the authors' expectations, the differences between boys and girls were larger in Germany ($d = 5.36$) than in the U.S. ($d = 1.82$). The observation that German boys threw much better than girls in the U.S. even though they do not throw much, have U.S.-like cultural support or encouragement for throwing, and throw as well as U.S. boys (Ehl et al. 2005) is consistent with our hypothesis and challenges the hypothesis that sex-biased training significantly contributes to the sex differences in throwing.

In summary, the evidence from a variety of disciplines is inconsistent with the hypothesis that environmental influences account for a large proportion of the sex differences in throwing. Nevertheless, the idea that sex differences in throwing result solely from environmental influences, including the idea that females are "conditioned by their sexist oppression in contemporary society" to be poor throwers (Young 2009:152) appears to

be the dominant view among scholars and journalists (e.g., Haspel 2012; Anthamatten 2014).

SEX DIFFERENCES IN THROWING BEHAVIOR IN SPORTS

Most hypotheses about the functions of sport agree that it likely had its origins as a way for males to develop and practice the skills necessary for primitive warfare and/or hunting (Sipes 1973; Chick et al. 1997; Carroll 2000; Lombardo 2012). Numerous studies demonstrate that boys consistently attain higher scores than girls on tests of motor proficiency skills that require the same kinds of bursts of power (e.g., jumping, sprinting, and throwing; Venetsanou and Kambas 2016) that are advantageous during fighting and hunting. The connection between primitive warfare and hunting techniques and sports that involve throwing, intercepting, and dodging projectiles is obvious. These observations are consistent with our hypothesis that throwing evolved as a male adaptation in the context of greater male than female participation in fighting and hunting. Because weapons are thrown much less frequently during modern combat or hunting than they were in the past, we predict that vestiges of the importance to males of developing throwing ability are present today in the prominence of throwing, intercepting, and dodging projectiles in sports.

Although sports are a human universal (Brown 1991) and despite the recent rapid increase in participation by females, especially in the U.S. (Shulman and Bowen 2001), they remain primarily a male endeavor (Guttmann 1991, 2004a,b). Deaner et al. (2012) confirmed the generalization that males not only play sports more often but are also more motivated to play sports than are females even when accounting for a country's gender inequality (Deaner et al. 2012, 2016; Deaner and Smith 2013; Balish et al. 2016). The generalizations that sports played by boys and men often either involve throwing projectiles for distance, velocity, or accuracy or avoiding getting hit by, catching, or hitting projectiles in clear parallel to the use of pro-

jectiles in primitive hunting and warfare (Deaner et al. 2012).

Boys in traditional societies throw projectile weapons during play more often than girls because boys play war and hunting games more often (Hoffman 1890; Eibl-Eibesfeldt 1989; Loy and Hesketh 1995; Chagnon 1997). For example, Himba boys in Namibia practice throwing stones for distance and accuracy more often than girls (Eibl-Eibesfeldt 1989) and Yanomamō boys start practicing with bows at a very early age (Chagnon 1997).

SEX DIFFERENCES IN THROWING DURING COMBAT

If throwing first evolved in the context of agonistic interactions, then we predict that throwing and projectile use would be male-biased and play a central role in primitive warfare. Primitive warriors used human muscle power to throw bolas (i.e., a number of balls connected by a cord that when thrown entangles the legs of prey), boomerangs, darts, knives, spears thrown with and without the aid of atlatls, sticks, and stones (Turney-High 1971; Oswalt 1976; van Creveld 1989). More advanced weapons (arrows, catapults, bullets, and missiles) rely on mechanical or chemical sources of energy to “throw” projectiles (van Creveld 1989; Crosby 2002). Although sticks and stones were undoubtedly the first thrown projectiles, they are used much less often by modern warriors. However, they remain effective weapons and are commonly used during conflicts with rivals, including against those with much more lethal weapons such as firearms. For example, modern rioters commonly throw stones, bricks, and bottles at police or soldiers armed with firearms. The threat of being injured or killed by thrown projectiles is treated as serious by targets who have been known to fire their guns at stone-throwing protestors (President’s Committee on Campus Unrest 1988).

In Australia, the bow was not yet invented so men and women used throwing sticks for defense before the introduction of firearms but, as elsewhere, men were much more likely than women to be warriors (Davidson 1936; Clarke 2003). Even though throwing is an important skill for Australian Aboriginal girls,

they remain inferior throwers compared to Australian Aboriginal boys. Aboriginal girls are even inferior throwers compared to boys in modern Germany where hunting and fighting by throwing is virtually nonexistent and sports that involve throwing are not popular (Thomas et al. 2010).

Collectively, these observations are consistent with our hypothesis that the causes of selection on throwing ability were stronger on males than on females and have resulted in the evolution of sex differences in throwing that are resistant to change by training. Nevertheless, environmental influences such as the different treatment of boys and girls by parents, teachers, and coaches during development and differential throwing and sports experience and opportunities due to cultural norms and expectations likely account for some of the sex differences in throwing (Cratty 1979; Young 1980; Sakurai and Miyashita 1983; Nelson et al. 1986; Thomas and Thomas 1988).

SEX DIFFERENCES IN THROWING BEHAVIOR DURING HUNTING

Throwing and the use of projectiles are common in primitive hunting, especially against large game (Oswalt 1976; van Creveld 1989; Churchill 1993; Hitchcock and Bleed 1997; Shea 2009). Hunting large game with projectile weapons is primarily a male activity (Murdock and Provost 1973; Marlowe 2007). Modern hunter-gatherer women may hunt but with the exception of the Agta of the Philippines (Goodman et al. 1985), they rarely use weapons capable of penetrating the flesh of their prey (Testart 1986 in Villotte and Knüsel 2014).

THE SEX DIFFERENCES IN ANATOMICAL FEATURES ASSOCIATED WITH THROWING ARE EVOLUTIONARY CONSEQUENCES OF SELECTION FAVORING MALE THROWING BIOLOGY

If the causes of selection are stronger on one sex, then the evolutionary consequences of that selection are more prominently expressed in that sex (Darwin 1871). The scientific literature is replete with examples of sex differences in anatomy and behavior

that have evolved as consequences of different causes of selection on males and females. An especially compelling example is the derived curvature (i.e., lumbar lordosis) and reinforcement of the lumbar vertebrae of human females to compensate for the ways that the expanded abdomen of pregnancy and enlarged breasts during lactation compromise locomotion by relocating the female's center of mass from over her pelvis to a position anterior to her hips (Whitcome et al. 2007; Masharawi et al. 2010; Hay et al. 2015). This female-only skeletal adaptation may have arisen with the advent of bipedalism in the Australopithecines prior to 2 million years ago (Whitcome et al. 2007). Males have not evolved these skeletal adaptations because the biomechanical challenges associated with being bipedal and pregnant or lactating are not causes of selection on male anatomy.

Given the biomechanical demands of the very rapid and coordinated movements of the humerus and forearm, trunk rotation, and striding toward the target during the mature throwing action (Wild 1938) and the historically important sex differences in throwing behavior during combat and hunting, we predict that selection has produced sex differences in anatomy associated with throwing.

If the consequences of throwing ability were stronger causes of selection on males than on females, then we predict sex differences to have evolved in the anatomical traits associated with throwing. However, intrasexual variation in these traits are also expected because bone morphology is affected by the interaction between the genetic information shaping it and the mechanical forces it is subjected to during development (Ruff 2003). Observations of little inter- or intrasexual variation in bone morphology are evidence of strong selection favoring a particular phenotype (Williams 1985; West-Eberhard 2003).

SEX DIFFERENCES IN THE PECTORAL GIRDLE

Anatomical sex differences in the pectoral girdle relevant to throwing performance are summarized in Table 3. Below we discuss how sexual dimorphism in each of these features

may contribute to the sex differences in throwing performance.

SHOULDER WIDTH

Shoulder width is a significant predictor of the torso's contribution to the kinetic energy of thrown projectiles (Maki 2013). That males, on average, have wider shoulders than females is well established (Stoudt et al. 1970; Malina and Zavaleta 1976; Lohman et al. 1988). Moreover, sex differences in shoulder width are large in both an Australian Aboriginal population in which both males and females throw in hunting and defense (Abbie 1957) and in 20th-century populations in the U.S. (Montoye et al. 1965) and England (Hiorns and Harrison 1985) in which throwing was likely much less important to survival. The large effect sizes in all of these studies indicate little overlap between men and women (Table 3). The sexual dimorphism in shoulder width develops during the pubertal growth spurt (Gasser et al. 2000) producing large sex differences in shoulder width in adolescents (Table 3). This sex difference suggests that there was strong selection on males beginning in puberty favoring arm and shoulder power production, both of which would contribute to throwing performance (Lundh et al. 2011) even if the primary cause of selection on arm and shoulder power production was success in hand-to-hand combat (Lombardo and Deaner 2016).

GLENOID FOSSAE

The articulations between the glenoid fossae of the scapulae and humeral heads form the shoulder joints. The orientation and size of the glenoid fossa affect throwing by affecting the range of motion (ROM) of the arms.

GLENOID INCLINATION

The upward/downward tilt of the glenoid fossa relative to the scapula is referred to as glenoid inclination (Figure 1). The cranial orientation of the glenoid fossae of other primates is inferred to be an adaptation for the overhead use of the arms during arm-hanging

TABLE 3
Sex differences in the anatomical characteristics of the pectoral girdle associated with throwing by modern humans

Anatomical characteristic	Condition that enhances throwing ability	Population time period	Population age group	Population identity and/or location (side measured)	Sample sizes M, F	Direction of sex difference	P	d	Reference
Shoulder width	Wider	20th century	Adults	Njalia Aborigines, Australia	22, 20	M > F	***	1.88	Abbie (1957)
		20th century	Adults	U.S.	418, 475	M > F	***	2.12	Montoye et al. (1965)
		20th century	Adolescents	England	61, 38	M > F	***	2.16	Hauspie et al. (1985)
		20th century	Adolescents	India	49, 38	M > F	***	1.89	Hauspie et al. (1985)
		20th century	Adults	England	132, 130	M > F	***	13.66	Hiorns and Harrison (1985)
Glenoid inclination	Lateral orientation	20th century	Adults	African and Euro-Americans (mean of right and left)	100, 72	Male more lateral than female	NS	0.14	Churchill et al. (2001)
		20th century	Adults	Japan (mean of right and left)	97, 108	M > F	***	0.33	Matsumura et al. (2014)
Glenoid version	Greater retroversion	20th century	Adults	African and Euro-Americans (mean of right and left)	100, 72	M > F	***	2.80	Churchill et al. (2001) ¹
Glenoid breadth	Wider	20th century	Adults	Guatemala (left)	65, 38	M > F	***	2.60	Frutos (2002)
		Pre-1800	Adults	New Zealand Polynesian (not indicated)	31, 37	M > F	***	2.53	Murphy (2002)
		10th and 11th centuries	Adults	East Anatolia (left)	47, 45	M > F	***	2.12	Özer et al. (2006)
		20th century	Adults	African and Euro-Americans (mean of right and left)	184, 184	M > F	***	2.04	Merrill et al. (2009) ¹

continued

TABLE 3
Continued

Glenoid height	Taller	20th century	Adults	African and Euro-Americans (left)	447, 277	M > F	***	1.98	Dabbs and Moore-Jansen (2010) ¹
		20th century 19th and 20th centuries	Adults	Black South Africans	60, 60	M > F	***	2.08	Macaluso (2011)
			Adults	Crete (left)	81, 66	M > F	***	2.51	Papaioannou et al. (2012)
		20th century	Adults	African and Euro-Americans (mean of right and left)	100, 72	M > F	***	2.61	Churchill et al. (2001) ¹
		20th century Pre-1800	Adults	Guatemala (left)	65, 37	M > F	***	2.50	Frutos (2002)
			Adults	New Zealand Polynesian (not indicated)	17, 22	M > F	***	2.55	Murphy (2002)
		10th and 11th centuries	Adults	East Anatolia (left)	45, 45	M > F	***	1.70	Özer et al. (2006)
		20th century	Adults	African and Euro-Americans (mean of right and left)	184, 184	M > F	***	1.25	Merrill et al. (2009) ¹
		20th century	Adults	African and Euro-Americans (left)	447, 277	M > F	***	2.35	Dabbs and Moore-Jansen (2010) ¹
Glenoid area	Larger	20th century	Adults	Black South Africans (left)	60, 60	M > F	***	2.27	Macaluso (2011)
		19th and 20th centuries	Adults	Crete (left)	81, 66	M > F	***	2.29	Papaioannou et al. (2012)
			Adults	Germany (not indicated)	114, 100	M > F	***	2.48	Prescher and Klümpen (1995)
		20th century	Adults	Black South Africans (left)	60, 60	M > F	***	2.37	Macaluso (2011)
Glenoid index = glenoid breadth/ height	Larger	20th century	Adults	Euro-Americans (mean of right and left)	46, 50	M < F	NS	0.20	Churchill and Trinkaus (1990)

continued

TABLE 3
Continued

Anatomical characteristic	Condition that enhances throwing ability	Population time period	Population age group	Population identity and/or location (side measured)	Sample sizes M, F	Direction of sex difference	P	d	Reference
Clavicle length	Longer	20th century	Adults	African and Euro-Americans (mean of right and left)	184, 184	M > F	***	1.40	Merrill et al. (2009) ¹
		20th century	Adults	Euro-Americans (right)	40, 39	M > F	NS	0.31	Churchill and Rhodes (2009)
		20th century	Adults	African Americans (right)	25, 25	M > F	NS	0.20	Churchill and Rhodes (2009)
		20th century	Adults	Aleutian Islanders (right)	24, 20	M > F	NS	0.19	Churchill and Rhodes (2009)
		20th century	Adults	African Americans (right)	50, 50	M < F	*	0.42	Terry (1932)
		20th century	Adults	Amritsar Zone, India (right)	236, 112	M > F	***	1.74	Jit and Singh (1966)
		20th century	Adults	Varanasi Zone, India (right)	97, 97	M > F	***	1.99	Singh and Gangrade (1968)
		20th century	Adults	Chandigarh Zone, India (right)	280, 80	M > F	***	1.82	Jit and Sahni (1983)
		Pre-European Settlement 20th century	Adults	New Zealand Polynesian (right)	12, 15	M > F	***	2.28	Murphy (1994)
		20th century	Adults	Chandigarh Zone, India (right)	748, 252	M > F	***	1.63	Kaur et al. (2002)
		21st century	Adults	Iran (left)	60, 60	M > F	***	1.77	Akhlaghi et al. (2012)
		21st century	Adults	Southern India (right)	185, 135	M > F	***	0.92	Nagarchi et al. (2014)
		21st century	Adults	North Karnataka Zone, India (right)	39, 30	M > F	***	0.87	Shobha et al. (2014)
		21st century	Adults	Spain (right)	45, 32	M > F	***	2.41	Alcina et al. (2015)

continued

TABLE 3
Continued

Humeral torsion—"twisting" of the humerus from medial orientation	Less (i.e., more retroversion—humerus head oriented posteriorly)	20th century	Adults	Sardinia (mean of right and left)	100, 99	M > F	NS	0.12	Gualdi-Russo (1998)
		Unknown	Adults	Euro-American (right)	17, 11	M < F	**	1.22	Edelson (1999)
		Unknown	Adults	African American (right)	15, 11	M < F	NS	0.68	Edelson (1999)
		Unknown	Adults	Alaskan Inuit (right)	15, 13	M < F	NS	0.87	Edelson (1999)
		Unknown	Adults	Northern Chinese (right)	16, 12	M < F	NS	0.12	Edelson (1999)
		Unknown	Adults	Bedouin (right)	14, 15	M < F	NS	0.73	Edelson (1999)
		Unknown	Adults	Native American—New Mexico (right)	16, 13	M < F	**	1.21	Edelson (1999)
		Unknown	Fetuses aged 5–9 months	African and Euro-American (right)	29, 21	M > F	NS	0.08	Edelson (1999)
		20th century	Adults	Euro-American (right)	26, 26	M < F	***	1.01	Rhodes and Churchill (2009)
		20th century	Adults	African American (right)	25, 25	M < F	**	0.74	Rhodes and Churchill (2009)
		Pre- and early post-Russian contact	Adults	Aleutian Islander (right)	25, 24	M < F	NS	0.29	Rhodes and Churchill (2009)
		21st century	Adults	Japan (mean of right and left)	97, 108	M < F	***	0.45	Matsumura et al. (2014)

See the text for complete descriptions of how the characteristics are related to throwing. To be included, sources of data needed to report the means, standard deviations, sample sizes, the time period when the individuals sampled lived, the age class of individuals, and the identity of the population to which individuals belonged. A *t*-test was used to statistically compare the sexes. NS = Not Significant.
* = $P \leq 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$. d = effect size (Cohen 1992).
*Samples overlap because investigators used samples in the Hamann-Todd Collection at the Cleveland Museum of Natural History, Cleveland, Ohio.

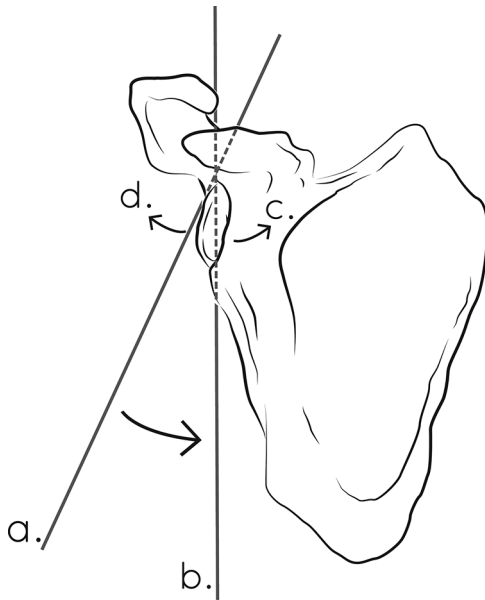


FIGURE 1. DORSAL VIEW OF LEFT SCAPULA INDICATING GLENOID FOSSA INCLINATION AND VERSION

a. Cranial inclination of the glenoid fossa. The cranial orientation of the glenoid fossa in other primates is inferred to be an adaptation for the overhead use of the arms during arm-hanging and climbing (Hunt 1991; Young et al. 2015). b. Lateral orientation of the glenoid fossa found in *Homo* facilitates throwing. c. Glenoid retroversion (i.e., dorsal tilt of the glenoid) facilitates throwing performance. d. Glenoid version (i.e., ventral tilt of the glenoid).

and climbing (Hunt 1991; Young et al. 2015). The Australopithecines had cranially oriented glenoid fossae leading to the inference that, although they were bipedal, they had not yet anatomically abandoned their ability to climb (Stern 2000; Green and Alemseged 2012; Larson 2012; Churchill et al. 2013). In contrast, members of the genus *Homo* have laterally oriented shoulder joints (Walker and Leakey 1993; Larson 2009; Young et al. 2015, but see Lordkipanidze et al. 2007) that facilitate throwing.

If the lateral orientation of the glenoid fossa facilitates high-speed throwing (Roach et al. 2013), they should, on average, be more laterally oriented in males than in females. Contrary to this prediction, there is no evidence of statistically significant sexual di-

morphism in this trait (Table 3). However, the trend of the sex difference is in our predicted direction (Table 3). For example, the glenoid fossae of adult African American and European American males, on average, were more laterally oriented than those of females (Churchill et al. 2001). The lack of statistically significant sexual dimorphism in glenoid inclination suggests two nonmutually exclusive explanations. First, there may be developmental constraints on adult glenoid inclination. Second, glenoid inclination, within a relatively wide range, may not affect throwing performance.

GLENOID VERSION

Glenoid version refers to the dorsal/ventral orientation of the glenoid surface and is defined as the angle between a line connecting the anterior and posterior rims of the glenoid and the line perpendicular to the scapular axis (Friedman et al. 1992; Figure 1). Glenoid fossae tilted ventrally relative to a lateral orientation are described as having positive version; those tilted dorsally are described as being retroverted. External rotation of the glenohumeral joint, and therefore throwing performance, improves with increased retroversion (Crockett et al. 2002; Whiteley 2007; Wyland et al. 2012; Roach et al. 2013). Glenoid version is influenced by activity and typically becomes retroverted with frequent throwing. For example, increased glenoid retroversion is commonly observed in the dominant shoulder of throwing athletes and is likely due to the mechanical stresses of throwing during skeletal development. Accordingly, glenoid retroversion is not well developed in the dominant shoulders of non-throwing athletes (Crockett et al. 2002; Wyland et al. 2012).

Matsumura et al. (2014) reported the glenoid fossae of males being significantly more retroverted, on average, than those of females but the effect size was moderate (Table 3). As expected, individuals with a history of playing throwing sports had greater glenoid retroversion in their dominant shoulders. High glenoid retroversion and low humeral torsion (see below) increase shoulder external rotation (Matsumura et al. 2014) and thus

they may be skeletal adaptations for overhand throwing.

GLENOID SHAPE

The glenoid fossa's size and shape affect throwing performance (Crockett et al. 2002; Wyland et al. 2012; Roach et al. 2013) because they influence the upper arm's ROM (Churchill and Trinkaus 1990). The glenoid fossa is pear-shaped (Merrill et al. 2009; Figure 2). Sexual dimorphisms in the breadth and height of the glenoid fossae of anatomically modern humans are large (Table 3), well known by anatomists (Prescher and Klümpen 1995; Merrill et al. 2009), and are used to identify the sex of the dead from skeletal remains (Di Vella et al. 1994; Frutos 2002; Murphy 2002). The wide glenoid fossae of modern humans may have resulted from selection on glenoid shape during a history of frequent overhand throwing (Churchill and Trinkaus 1990). Male shoulders have larger joint surfaces because their glenoid fossae area is larger than that of females (Table 3). The larger joint surfaces relative to body stature in males are likely related to greater male muscularity and skeletal robusticity (Churchill and Trinkaus 1990). The dimensions of the glenoid fossae develop early in fetal life (Hrdlička 1942) and the sex differences are apparent by 4.2 years of age (Humphrey 1998) suggesting that selection shaped these sex differences.

The ratio between the glenoid's greatest breadth and height, glenoid index (Figure 2), is related to the degree of dorsoventral glenohumeral movement (Churchill and Trinkaus 1990). The glenoid index of Neanderthals was less than that of modern humans. This likely limited the ROM of Neanderthal upper arms and consequently their ability to throw projectiles with an overhand motion (Churchill and Trinkaus 1990). Modern males typically have larger glenoid indices than females, but the difference is not always statistically significant and effect sizes range from small to moderate (Table 3). The sexual dimorphism of the glenoid is not due solely to sex differences in body size; when the ratios between humerus length and either glenoid

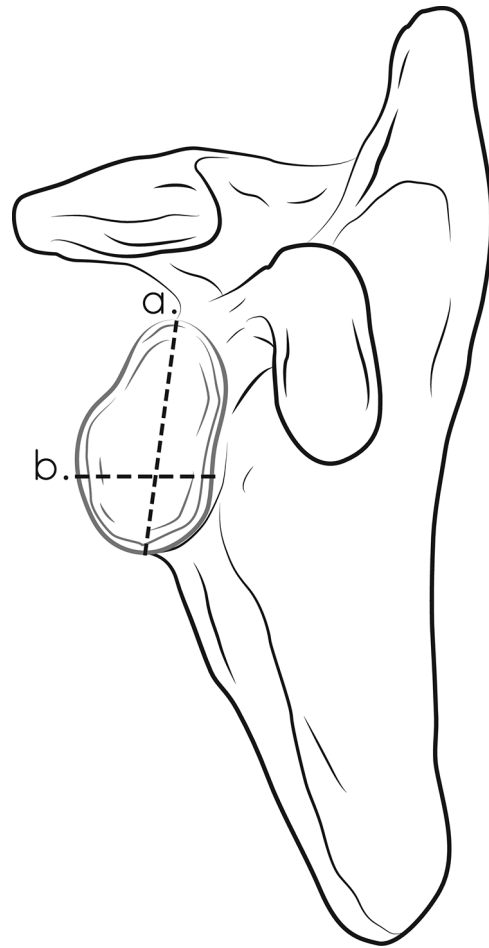


FIGURE 2. GLENOID FOSSA SHAPE

Glenoid index = (b/a) ; a larger glenoid index and glenoid area facilitate throwing by increasing the range of motion of the glenohumeral joint.

breadth or height are used to control for body size, male glenoid fossae are larger (Churchill and Trinkaus 1990). Consequently, resulting greater ROM in the glenohumeral joints of males likely contributes to the male advantage in throwing speed.

Interestingly, the dimensions of the glenoid, especially the lower glenoid where it articulates with the humeral head, show the least individual variation of all of the anatomical structures of the scapula (Hrdlička 1942). This suggests that the ROM requirements of the overhead arm actions needed for throw-

ing, and possibly overhand clubbing (Young 2009), were important causes of selection on glenoid shape.

CLAVICLE LENGTH

The clavicle anchors the shoulder complex to the torso and its length may influence throwing performance (Larson 2007b, 2015; Roach and Richmond 2015). Short clavicles are thought to result in poor throwing ability mostly because individuals with pathologically short clavicles report an inability to throw well (Larson 2007b). If clavicle length is directly related to throwing ability, we predict that the clavicle lengths of males would be more advantageous for throwing than those found in females. Indeed, males have longer clavicles, on average, than females and the difference between the sexes is large (Table 3) and is used to identify the sex of the dead from skeletal remains (Terry 1932; Singh and Gangrade 1968; Murphy 1994; Frutos 2002; Kaur et al. 2002; Özer et al. 2006; Akhlaghi et al. 2012; Papaioannou et al. 2012; Alcina et al. 2015). However, the sex difference in clavicle length is related to larger male body size. In one study, relative clavicle length was not related to throwing performance in Daasanach men in East Africa who are expert throwers and frequently use throwing sticks during hunting (Roach and Richmond 2015).

HUMERAL TORSION

A central assumption in the clinical and sports medicine literature is that the normal condition of the shoulder joint is a laterally facing glenoid fossa with a medially facing humeral head (Larson 2007a). In reality, the humeral head is slightly twisted and thus does not often face directly medially. This condition is referred to as humeral torsion and is measured as the angle between a line bisecting the proximal head of the humerus and one between the lateral and medial epicondyles of the distal humerus (Figure 3; Krahl and Evans 1945). Both heredity and the muscular forces acting on the upper arm and shoulder joint during development affect the degree of torsion (Evans and Krahl

1945). However, once growth has stopped, increased training does not alter the degree of torsion (Meyer et al. 2011). Low humeral torsion (Figure 4) is typically described as a skeletal adaptation to frequent, excessive external rotation from overhand throwing during growth (Pieper 1998).

Low humeral torsion is advantageous for throwing because it enables throwers to increase the cocking phase of their throw producing increased projectile velocity (Rhodes and Knüsel 2005; Roach et al. 2012). Observations of low humeral torsion in modern athletes playing sports requiring frequent overhand throwing are common (Crockett et al. 2002; Osbahr et al. 2002; Whiteley et al. 2009). From these observations, Roach et al. (2013) hypothesized that *Homo erectus*, because of its low humeral torsion, was capable of high-speed overhand throwing. Early hominins were characterized by low humeral torsion (Walker and Leakey 1993; Larson et al. 2007; Lordkipanidze et al. 2007; Lovejoy et al. 2009; Churchill et al. 2013). However, Larson (2015) cogently argued that it is difficult to reconcile the inference of the growing importance of throwing to the survival of post-*H. erectus* hominins with the subsequent increase rather than expected decrease in humeral torsion. Nevertheless, because low humeral torsion facilitates high-speed throwing, we predict that males exhibit lower humeral torsion than females.

Consistent with our prediction, humeral torsion is (with few exceptions) lower in males, on average, than in females (Table 3). However, the extensive variation in humeral torsion angles between dominant and nondominant arms within individuals (e.g., throwing arms show significantly less torsion than nonthrowing arms in baseball pitchers; Pieper 1998; Whiteley et al. 2009), between sexes, and within and among populations (Table 3) suggests three, nonmutually exclusive explanations. First, within a fairly wide range, the actual angle of humeral torsion may have only a modest direct effect on throwing ability. Therefore, humeral torsion angle may not have been an important target of selection during the evolution of throwing in hominins (Maki 2013). Second, because frequent throwing decreases humeral torsion (Krahl

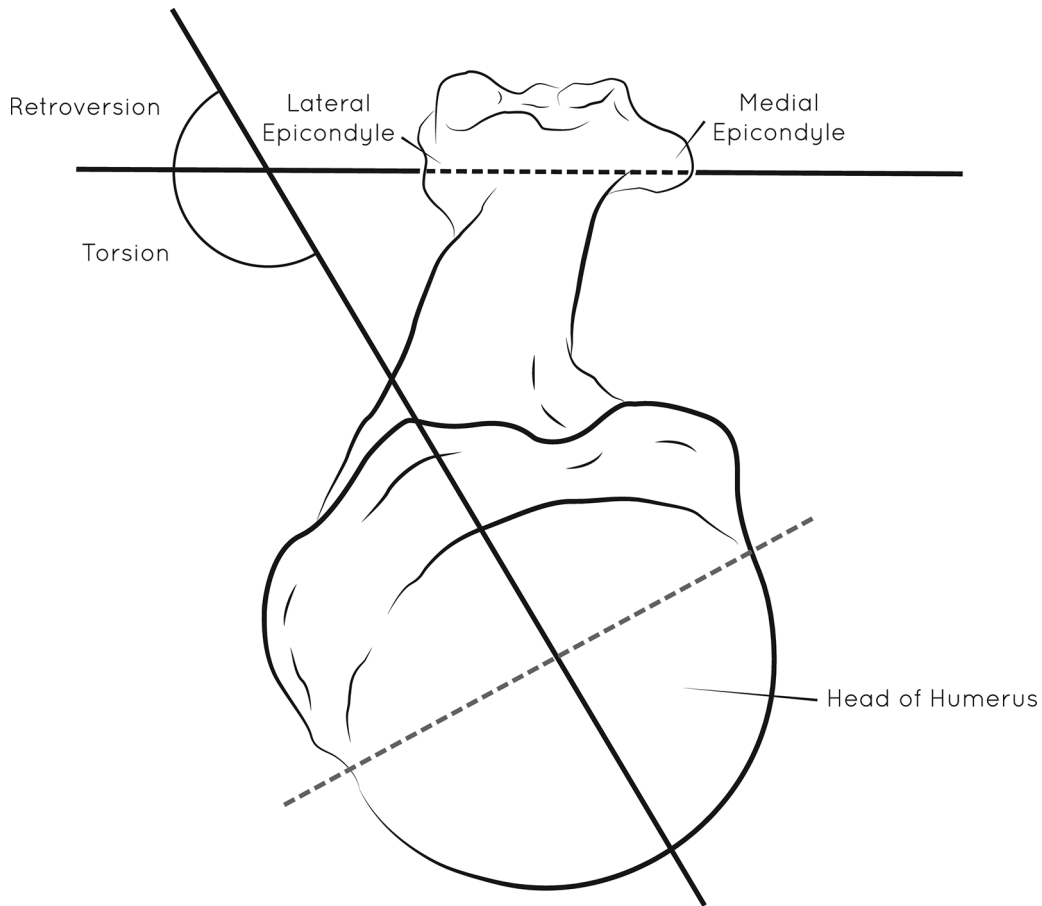


FIGURE 3. TORSION OF LEFT HUMERUS

Humeral torsion is measured as the angle between a line bisecting the proximal head of the humerus and one between the lateral and medial epicondyles of the distal humerus (Krahl and Evans 1945).

and Evans 1945; Edelson 1999; Rhodes 2006; Cowgill 2007), especially in dominant arms, the relatively wide range of humeral torsion angles observed in modern humans suggests there are population, sex, and individual differences in activities requiring external rotation of the arm (Cowgill 2007). Third, we hypothesize that the wide variation in humeral torsion reflects evidence of relaxed selection on this trait during the evolution of anatomically modern humans. The transition from hunter-gatherer to agricultural subsistence economies may have relaxed selection on throwing adaptations in humeral torsion. For example, male Native American hunter-gatherers in southern California from

groups that used spears and atlatls displayed greater bilateral asymmetry in humeral torsion than did males in contemporaneous groups that were agriculturalists (Gjerdrum et al. 2003). This pattern suggests important differences between these populations in the frequency of overhand motions such as throwing.

SUMMARY OF THE SEX DIFFERENCES IN THE PECTORAL GIRDLE RELATED TO THE SEX DIFFERENCES IN THROWING

The sex differences in the anatomy of the pectoral girdle are, with few exceptions, in the direction that would result in superior



FIGURE 4. VIEW OF RIGHT HUMERUS COMPARING A HUMERUS SHOWING NO TORSION AND ONE WITH INCREASED TORSION

Low humeral torsion is advantageous for throwing because it increases the cocking phase of the throw resulting in increased projectile velocity (Rhodes and Knüsel 2005; Roach et al. 2012).

throwing performance by males. Moreover, they are consistent with the nonmutually exclusive hypotheses that overhand motions during growth are performed more often by males than females and shape the pectoral girdle in ways that result in superior male throwing performance and throwing ability was a more important cause of selection on male than female throwing ability. Whether the skeletal structures of male pectoral girdles are more responsive to frequent overhand throwing than those of females remains an open question.

OTHER SEX DIFFERENCES IN ANATOMY AFFECTING THROWING ABILITY

Several anatomical sex differences that are not specific to throwing contribute to male

throwing superiority. Bodies that are taller, have more muscle mass, and longer limbs are capable of generating more muscular torque and power during throwing (Chu et al. 2009). Some of these features may have been targets of selection for reasons (e.g., climatic adaptations; Ruff 1991, 1994) other than throwing but once evolved would have contributed to the sex differences in throwing performance. Thus, some traits may be considered to be exaptations for throwing (Maki 2013). Therefore, regardless of the multiple synergistic causes of selection that favored larger males than females during the course of human evolution, sex differences in various components of body size have produced superior male throwing performance. As these traits evolved and produced superior throwing performance, they also may have reinforced the adoption of throwing weapons by *H. erectus* and its successors.

SEX DIFFERENCES IN BODY SIZE

Males, on average, are larger than females. Van den Tillaar and Ettema (2004) studied the relationship between sex, body size, isometric strength, and throwing velocity in Norwegian team handball players and concluded that fat-free body mass, independent of sex, is the factor that contributes the most to throwing velocity. From that result, they argued that the greater muscle mass of males is the primary cause of the sex difference in throwing velocity. Within sexes, larger athletes typically throw faster (Pyne et al. 2006) and farther (Morrow et al. 1982).

SEX DIFFERENCES IN STRENGTH

Muscular strength is positively correlated with the ability to throw fast and far. The muscular strength differences between males and females are well established (Bishop et al. 1987; Lassek and Gaulin 2009). For example, female strength even when controlling for differences in fat-free body mass, limb, or body size ranges from 40–75% of male strength (Bishop et al. 1987). The upper-body strength of males (Bishop et al. 1987; Lassek

and Gaulin 2009) contributes to the sex differences in throwing velocity.

compromises female ability to throw by inhibiting torso rotation.

SEX DIFFERENCES IN ARM LENGTH

Longer arms produce greater external arm rotation during throwing resulting in faster throws (Roach et al. 2013). Males have longer upper arms and forearms than females (Gindhart 1973; Dittrick and Suchey 1986; Steyn and İşcan 1997; İşcan et al. 1998; Kranioti and Michalodimitrakis 2009; Wyland et al. 2012; Waidhofer and Kirchengast 2015). Arm bone lengths can be used to identify the sex of the dead from skeletal remains (Krogman 1962; Stewart 1979; Bass 1995). Moreover, males have relatively longer forearms than females (Fullenkamp et al. 2008) and this difference emerges in utero (Tanner 1978). The sex difference in forearm development is consistent with a hypothesis of strong selection on male forearm length for throwing because disproportionately long forearms increase the extent of arm external rotation thereby increasing throwing velocity.

SEX DIFFERENCES IN HAND AND FINGER LENGTH

The longer hands and fingers of males at the ends of their longer arms (Peters et al. 2002) contribute to greater external arm rotation and thus greater throwing speed. Young (2009) argued that this advantage during overhand clubbing in hunting and combat may have favored the evolution of longer arms, hands, and fingers in males than in females.

SEX DIFFERENCES IN WAIST LENGTH

A tall, mobile waist decouples the hips and thorax during throwing permitting more torso rotation to produce the high torque needed to load the shoulder's elastic elements and produce faster throws (Roach et al. 2013). When height is controlled for, anatomically modern males have taller waists than females (Fullenkamp et al. 2008) facilitating faster throwing by males. It is not yet known whether the greater lumbar lordosis of females (Whitcome et al. 2007; Hay et al. 2015)

SEX DIFFERENCES IN LEG LENGTH

Regardless of the primary causes of selection that shaped the sexual dimorphism in leg length, longer legs help males throw faster and farther than females (Maki 2013). Males have longer legs, on average, than females and femur length is used to identify the sex of the dead from skeletal remains (Krogman 1962; Stewart 1979; Bass 1995). Moreover, leg length is important because stride length during overhand throwing contributes nearly 70% of the variation in throwing velocity (Stodden et al. 2006a). The overhand throwing stride length, as measured as percent of height, of elite female baseball pitchers (approximately 70%) was significantly shorter than that of elite male pitchers (approximately 78%; Chu et al. 2009).

SEX DIFFERENCES IN INJURIES ASSOCIATED WITH THROWING

If throwing was primarily a male activity, then we predict the existence of sex differences in injuries associated with throwing. For example, lesions of tendon attachments (i.e., enthesopathies) of the medial epicondyle were more common in males than females in prehistoric, preindustrial historic, and modern European populations (Villotte and Knüsel 2014). These injuries are evidence indicative of a sexual division of labor in those populations in arm movements that involve external rotation of the arm like those used in throwing, overhand clubbing, or in the use of an adze or axe. Sex differences in extensive training in these activities during growth may have contributed to the greater frequency of male medial epicondyle enthesopathies (Villotte and Knüsel 2014).

Although there are few sex differences in injury rates in college athletes in the U.S., females are more likely than males to injure their shoulders in swimming and water polo (Sallis et al. 2001), sports that require repetitive overhand shoulder motions, including external and internal rotation, against water

resistance. Water polo also requires throwing. The greater frequency of shoulder injuries in females suggests a history of stronger selection on males than females for repetitive forceful movements of their shoulders such as those that would be caused by the frequent throwing of weapons. Similarly, during ergonomic studies females had a higher prevalence of upper musculoskeletal disorders associated with repetitive motions (Nordander et al. 2009). Together, these injury data are consistent with our hypothesis because they are indirect evidence that males have evolved greater resistance than females to injuries caused by repetitive forceful movements of their upper extremities such as those associated with throwing.

CONCLUDING REMARKS

Sex differences in human behavior are subjects of notoriously contentious debates. However, there can be little doubt that there are innate sex differences in strength and speed. Throwing must be now added to that list.

We do not dispute the observation that *all* modern humans have the basic adaptations for throwing. Our hypothesis is that males are better adapted for throwing than females because during human evolution the consequences of throwing success or failure on fitness have been a stronger cause of selection on males. To be clear, we are not arguing that socialization has no effect on the sex differences in throwing or that females cannot become effective throwers. In fact, we presented evidence that training can improve female throwing performance and that some females are expert throwers. The evi-

dence that socialization plays a role in producing sex differences is not incompatible with the evolutionary approach we take here. The patterns of sex differences in a variety of traits often reflect differences in cultural norms and practices and are related to adult reproductive strategies and the causes of selection that shaped those strategies (Low 2000). Therefore, from an evolutionary perspective it is extremely unlikely that socialization alone can explain all of the sex differences in throwing. We have presented several lines of evidence that support our hypothesis that sex differences in throwing have a partly innate basis. Moreover, the evidence is inconsistent with the view held by many that sex differences in throwing are due solely to environmental causes (Table 1). In our view, the consilience of multiple, independent lines of evidence from anatomy, anthropology, archeology, childhood development, ethology, primatology, and sports sciences that were collected for purposes other than testing our hypothesis strengthens its support (Wilson 1998).

ACKNOWLEDGMENTS

We thank N. MacDonald, T. Strickler, and P. Thorpe for comments on earlier versions of the paper. MacDonald provided useful comments that helped us clarify some of our arguments. Strickler helped clarify some questions we had about anatomy. Thorpe suggested that we look at youth track and field performances and alerted us to the literature on the evolution of female lumbar lordosis. C. Shaffer alerted us to an importance source of information. K. Sanders drew the illustrations. We were supported by the Biology and Psychology departments at Grand Valley State University and a GVSU Center for Scholarly and Creative Excellence Mini-Grant.

REFERENCES

- Abbie A. A. 1957. Metrical characters of a central Australian tribe. *Oceania* 27:220–243.
- Akhlaghi M., Moradi B., Hajibeygi M. 2012. Sex determination using anthropometric dimensions of the clavicle in Iranian population. *Journal of Forensic and Legal Medicine* 19:381–385.
- Alcina M., Rissech C., Clavero A., Turbón D. 2015. Sexual dimorphism of the clavicle in a modern Spanish sample. *European Journal of Anatomy* 19:73–83.
- Anthamatten E. 2014. What does it mean to “throw like a girl?” *New York Times* 24 August. <http://opinionator.blogs.nytimes.com/2014/08/24/what-does-it-mean-to-throw-like-a-girl/>.
- Axe M. J., Snyder-Mackler L., Konin J. G., Strube M. J. 1996. Development of a distance-based interval throwing program for Little League-aged athletes. *American Journal of Sports Medicine* 24:594–602.
- Axe M. J., Strube M., Osinski D., Andrews J. R., Snyder-Mackler L. 2014. A speed distance-based classification system for injury prevention and research in international and domestic youth baseball players. *International Journal of Sports Physical Therapy* 9:346–355.

- Balish S. M., Deaner R. O., Rainham D., Blanchard C. 2016. Sex differences in sport remain when accounting for countries' gender inequality. *Cross-Cultural Research* 50:395–414.
- Bass W. M. 1995. *Human Osteology: A Laboratory and Field Manual*. Fourth Edition. Columbia (Missouri): Missouri Archaeological Society.
- Benenson J. F., Tennyson R., Wrangham R. W. 2011. Male more than female infants imitate propulsive motion. *Cognition* 121:262–267.
- Berenbaum S. A. 2011. The importance of puberty in adolescent development. Pages 95–104 in *Bio-social Foundations of Family Processes*, edited by A. Booth, S. M. McHale, and N. S. Landale. New York: Springer.
- Bingham P. M., Souza J. 2009. *Death From a Distance and the Birth of a Humane Universe: Human Evolution, Behavior, History, and Your Future*. Scotts Valley (California): BookSurge Publishing.
- Bingham P. M., Souza J. 2010. Our Response to the Authors' Objections. <http://www.deathfromadistance.com/posts/common-misunderstandings/2010/05/objections-to-our-approach-in-the-evolution-of-human-language/#more-38>.
- Bishop P., Cureton K., Collins M. 1987. Sex difference in muscular strength in equally-trained men and women. *Ergonomics* 30:675–687.
- Brown D. E. 1991. *Human Universals*. New York: McGraw-Hill.
- Browne K. 2007. *Co-Ed Combat: The New Evidence that Women Shouldn't Fight the Nation's Wars*. New York: Sentinel.
- Butterfield S. A., Loois E. M. 1993. Influence of age, sex, balance, and sports participation on development of throwing by children in grades K–8. *Perceptual and Motor Skills* 76:459–464.
- Calvin W. H. 1983. *The Throwing Madonna: Essays on the Brain*. New York: McGraw-Hill.
- Carroll D. M. 2000. *An Interdisciplinary Study of Sports as a Symbolic Hunt: A Theory of the Origin and Nature of Sport Based on Paleolithic Hunting*. Lewiston (New York): Edwin Mellen Press.
- Ceci S. J., Williams W. M. 2011. Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences of the United States of America* 108:3157–3162.
- Chagnon N. A. 1988. Life histories, blood revenge, and warfare in a tribal population. *Science* 239:985–992.
- Chagnon N. A. 1997. *Yanomamö*. Fifth Edition. New York: Harcourt Brace College Publishers.
- Chick G., Loy J. W., Miracle A. W. 1997. Combative sport and warfare: a reappraisal of the spillover and catharsis hypotheses. *Cross-Cultural Research* 31:249–267.
- Chow R. S., Medri M. K., Martin D. C., Leekam R. N., Agur A. M., McKee N. H. 2000. Sonographic studies of human soleus and gastrocnemius muscle architecture: gender variability. *European Journal of Applied Physiology* 82:236–244.
- Chu Y., Fleisig G. S., Simpson K. J., Andrews J. R. 2009. Biomechanical comparison between elite female and male baseball pitchers. *Journal of Applied Biomechanics* 25:22–31.
- Churchill R. S., Brems J. J., Kotschi H. 2001. Glenoid size, inclination, and version: an anatomic study. *Journal of Shoulder and Elbow Surgery* 10:327–332.
- Churchill S. E. 1993. Weapon technology, prey size selection, and hunting methods in modern hunter-gatherers: implications for hunting in the Palaeolithic and Mesolithic. *Archaeological Papers of the American Anthropological Association* 4:11–24.
- Churchill S. E., Rhodes J. A. 2009. The evolution of the human capacity for “killing at a distance”: the human fossil evidence for the evolution of projectile weaponry. Pages 201–210 in *The Evolution of Hominin Diets: Integrating Approaches to the Study of Palaeolithic Subsistence*, edited by J.-J. Hublin and M. P. Richards. Dordrecht (The Netherlands): Springer.
- Churchill S. E., Trinkaus E. 1990. Neandertal scapular glenoid morphology. *American Journal of Physical Anthropology* 83:147–160.
- Churchill S. E., Holliday T. W., Carlson K. J., Jashashvili T., Macias M. E., Mathews S., Sparling T. L., Schmid P., de Ruiter D. J., Berger L. R. 2013. The upper limb of *Australopithecus sediba*. *Science* 340:1233477.
- Clarke P. 2003. *Where the Ancestors Walked: Australia as an Aboriginal Landscape*. Crows Nest (Australia): Allen and Unwin.
- Cohen J. 1992. A power primer. *Psychological Bulletin* 112:155–159.
- Cohen-Bendahan C. C. C., van de Beek C., Berenbaum S. A. 2005. Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. *Neuroscience & Biobehavioral Reviews* 29:353–384.
- Cowgill L. W. 2007. Humeral torsion revisited: a functional and ontogenetic model for populational variation. *American Journal of Physical Anthropology* 134:472–480.
- Cratty B. J. 1979. *Perceptual and Motor Development in Infants and Children*. Second Edition. Englewood Cliffs (New Jersey): Prentice-Hall.
- Crockett H. C., Gross L. B., Wilk K. E., Schwartz M. L., Reed J., O'Mara J., Reilly M. T., Dugas J. R., Meister K., Lyman S., Andrews J. R. 2002. Osseous adaptation and range of motion at the glenohumeral joint in professional baseball pitchers. *American Journal of Sports Medicine* 30:20–26.
- Crosby A. W. 2002. *Throwing Fire: A History of Projectile Technology*. New York: Cambridge University Press.
- Dabbs G. R., Moore-Jansen P. H. 2010. A method for estimating sex using metric analysis of the scapula. *Journal of Forensic Sciences* 55:149–152.
- Darlington P. J. 1975. Group selection, altruism, reinforcement, and throwing in human evolution. *Pro-*

- ceedings of the National Academy of Sciences of the United States of America* 72:3748–3752.
- Darwin C. 1871. *The Descent of Man, and Selection in Relation to Sex*. London (United Kingdom): John Murray.
- Davidson D. S. 1936. Australian throwing-sticks, throwing-clubs, and boomerangs. *American Anthropologist* 38: 76–100.
- Deaner R. O., Smith B. A. 2013. Sex differences in sports across 50 societies. *Cross-Cultural Research* 47: 268–309.
- Deaner R. O., Geary D. C., Puts D. A., Ham S. A., Kruger J., Fles E., Winegard B., Grandis T. 2012. A sex difference in the predisposition for physical competition: males play sports much more than females even in contemporary U.S. *PLOS ONE* 7:e49168.
- Deaner R. O., Balish S. M., Lombardo M. P. 2016. Sex differences in sports interest and motivation: an evolutionary perspective. *Evolutionary Behavioral Sciences* 10:73–97.
- Department of the Army. 2005. *Field Manual 3-23.30: Grenades and Pyrotechnic Signals*. Washington (DC): Department of the Army.
- Di Vella G., Campobasso C. P., Dragone M., Introna F. Jr. 1994. Skeletal sex determination by scapular measurements. *Bollettino della Società Italiana di Biologia Sperimentale* 70:299–305.
- Dittrick J., Suchey J. M. 1986. Sex determination of prehistoric central California skeletal remains using discriminant analysis of the femur and humerus. *American Journal of Physical Anthropology* 70:3–9.
- Downey G. 2010. Throwing like a Brazilian: on ineptness and a skill-shaped body. Pages 297–326 in *The Anthropology of Sport and Human Movement: A Biocultural Perspective*, edited by R. R. Sands and L. R. Sands. Lanham (Maryland): Lexington Books.
- Duffy L. J., Baluch B., Ericsson K. A. 2004. Dart performance as a function of facets of practice amongst professional and amateur men and women players. *International Journal of Sports Psychology* 35:232–245.
- Dunsworth H., Challis J. H., Walker A. 2003. Throwing and bipedalism: a new look at an old idea. Pages 105–110 in *Walking Upright*, edited by J. L. Franzen, M. Köhler, and S. Moya-Sola. Frankfurt (Germany): Senckenberg Institute.
- Edelson G. 1999. Variations in the retroversion of the humeral head. *Journal of Shoulder and Elbow Surgery* 8:142–145.
- Ehl T., Robertson M. A., Langendorfer S. J. 2005. Does the throwing “gender gap” occur in Germany? *Research Quarterly for Exercise and Sport* 76:488–493.
- Eibl-Eibesfeldt I. 1989. *Human Ethology*. New York: Aldine De Gruyter.
- Ellis L., Hershberger S., Field E., Wersinger S., Pellis S., Geary D., Palmer C., Hoyenga K., Hetsroni A., Karadi K. 2008. *Sex Differences: Summarizing More than a Century of Scientific Research*. New York: Psychology Press.
- Espenschade A. 1960. Motor development. Pages 419–439 in *Science and Medicine of Exercise and Sports*, edited by W. R. Johnson. New York: Harper.
- Evaldsson A.-C. 2003. Throwing like a girl?: situating gender differences in physicality across game contexts. *Childhood* 10:475–497.
- Evans F. G., Krahll V. E. 1945. The torsion of the humerus: a phylogenetic survey from fish to man. *American Journal of Anatomy* 76:303–337.
- Fallows J. 1996. Throwing like a girl. *The Atlantic* 278: 84–87.
- Fisher R. A. 1958. *The Genetical Theory of Natural Selection*. Second Revised Edition. New York: Dover Publications.
- Fredrickson B. L., Harrison K. 2005. Throwing like a girl: self-objectification predicts adolescent girls’ motor performance. *Journal of Sport and Social Issues* 29: 79–101.
- Friedman R. J., Hawthorne K. B., Genev B. M. 1992. The use of computerized tomography in the measurement of glenoid version. *Journal of Bone and Joint Surgery* 74:1032–1037.
- Frutos L. R. 2002. Determination of sex from the clavicle and scapula in a Guatemalan contemporary rural indigenous population. *American Journal of Forensic Medicine and Pathology* 23:284–288.
- Fullenkamp A. M., Robinette K. M., Daanen H. A. M. 2008. Gender differences in NATO anthropometry and the implication for protective equipment. Wright-Patterson AFB (Ohio): Air Force Research Laboratory. <http://www.dtic.mil/get-tr-doc/pdf?AD=ADA491083>.
- Gasser T., Sheehy A., Molinari L., Largo R. H. 2000. Sex dimorphism in growth. *Annals of Human Biology* 27:187–197.
- Geary D. C. 1999. Evolution and developmental sex differences. *Current Directions in Psychological Science* 8:115–120.
- Gindhart P. S. 1973. Growth standards for the tibia and radius in children aged one month through eighteen years. *American Journal of Physical Anthropology* 39:41–48.
- Gjerdrum T., Walker P. L., Andrushko V. 2003. Humeral retroversion: an activity pattern index in prehistoric southern California. *American Journal of Physical Anthropology* 120(Supplement 36):100–101.
- Glenday C. 2007. *Guinness World Records 2008*. New York: Bantam Books.
- Goldstein J. 1995. Aggressive toy play. Pages 127–147 in *The Future of Play Theory: A Multidisciplinary Inquiry Into the Contributions of Brian Sutton-Smith*, edited by A. D. Pellegrini. Albany (New York): State University of New York Press.
- Gómez J. M., Verdú M., González-Megías A., Méndez M. 2016. The phylogenetic roots of human lethal violence. *Nature* 538:233–237.

- Goodall J. 1986. *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge (Massachusetts): Belknap Press of Harvard University Press.
- Goodman M. J., Griffin P. B., Estioko-Griffin A. A., Grove J. S. 1985. The compatibility of hunting and mothering among the Agta hunter-gatherers of the Philippines. *Sex Roles* 12:1199–1209.
- Green D. J., Alemseged Z. 2012. *Australopithecus afarensis* scapular ontogeny, function, and the role of climbing in human evolution. *Science* 338:514–517.
- Gualdi-Russo E. 1998. Study on long bones: variation in angular traits with sex, age, and laterality. *Anthropologischer Anzeiger* 56:289–299.
- Gurven M., Hill K. 2009. Why do men hunt? A reevaluation of “man the hunter” and the sexual division of labor. *Current Anthropology* 50:51–74.
- Guttmann A. 1991. *Women's Sports: A History*. New York: Columbia University Press.
- Guttmann A. 2004a. *From Ritual to Record: The Nature of Modern Sports*. New York: Columbia University Press.
- Guttmann A. 2004b. *Sports: The First Five Millennia*. Amherst (Massachusetts): University of Massachusetts Press.
- Haspel T. 2012. Throw like a girl? With some practice, you can do better. *Washington Post* 10 September. https://www.washingtonpost.com/national/health-science/throw-like-a-girl-with-some-practice-you-can-do-better/2012/09/10/9ffc8bc8-dc09-11e1-9974-5c975ae4810f_story.html.
- Haubenstricker J. L., Seefeldt V. 1986. Acquisition of motor skills during childhood. Pages 41–102 in *Physical Activity and Well-Being*, edited by V. Seefeldt. Reston (Virginia): American Alliance for Health, Physical Education, Recreation and Dance.
- Hauspie R., Das S. R., Preece M. A., Tanner J. M., Susanne C. 1985. Decomposition of sexual dimorphism in adult size of height, sitting height, shoulder width and hip width in a British and West Bengal sample. Pages 207–215 in *Human Sexual Dimorphism*, edited by J. Ghesquiere, R. D. Martin, and F. Newcombe. Philadelphia (Pennsylvania): Taylor & Francis.
- Hay O., Dar G., Abbas J., Stein D., May H., Masharawi Y., Peled N., HersHKovitz I. 2015. The lumbar lordosis in males and females, revisited. *PLOS ONE* 10:0133685.
- Hiorns R. W., Harrison G. A. 1985. Sex differences in anthropometric, endocrinological and behavioural traits in a group of Oxfordshire villages. Pages 217–221 in *Human Sexual Dimorphism*, edited by J. Ghesquiere, R. D. Martin, and F. Newcombe. Philadelphia (Pennsylvania): Taylor & Francis.
- Hitchcock R., Bleed P. 1997. Each according to need and fashion. Pages 345–368 in *Projectile Technology*, edited by H. Knecht. New York: Plenum Press.
- Hoffman W. J. 1890. Remarks on Ojibwa ball play. *American Anthropologist* A3:133–136.
- Hrdlička A. 1942. The adult scapula. Additional observations and measurements. *American Journal of Physical Anthropology* 29:363–415.
- Humphrey L. T. 1998. Growth patterns in the modern human skeleton. *American Journal of Physical Anthropology* 105:57–72.
- Hunt K. D. 1991. Mechanical implications of chimpanzee positional behavior. *American Journal of Physical Anthropology* 86:521–536.
- Hyde J. S. 2005. The gender similarities hypothesis. *American Psychologist* 60:581–592.
- Isaac B. 1987. Throwing and human evolution. *African Archaeological Review* 5:3–17.
- Isaac B., Isaac G. 2011. Unexpected trajectories: a history of Niuean throwing stones. *Journal of the Polynesian Society* 120:369–402.
- İşcan M. Y., Loth S. R., King C. A., Shihai D., Yoshino M. 1998. Sexual dimorphism in the humerus: a comparative analysis of Chinese, Japanese and Thais. *Forensic Science International* 98:17–29.
- Janssen I., Heymsfield S. B., Wang Z., Ross R. 2000. Skeletal muscle mass and distribution in 468 men and women aged 18–88 yr. *Journal of Applied Physiology* 89:81–88.
- Jardine R., Martin N. G. 1983. Spatial ability and throwing accuracy. *Behavior Genetics* 13:331–340.
- Jit I., Sahni D. 1983. Sexing the north Indian clavicles. *Journal of Anatomical Society of India* 32:61–72.
- Jit I., Singh S. 1966. The sexing of the adult clavicles. *Indian Journal of Medical Research* 54:551–571.
- Kaur H., Harjeet, Sahni D., Jit I. 2002. Length and curves of the clavicle of northwest Indians. *Journal of the Anatomical Society of India* 51:199–209.
- Keeley L. H. 1996. *War Before Civilization*. New York: Oxford University Press.
- Kimura D. 1999. *Sex and Cognition*. Cambridge (Massachusetts): MIT Press.
- Knüsel C. J. 1992. The throwing hypothesis and hominid origins. *Human Evolution* 7:1–7.
- Krahl V. E., Evans F. G. 1945. Humeral torsion in man. *American Journal of Physical Anthropology* 3:229–253.
- Kranioti E. F., Michalodimitrakis M. 2009. Sexual dimorphism of the humerus in contemporary Cretans—a population-specific study and a review of the literature. *Journal of Forensic Sciences* 54:996–1000.
- Krogman W. M. 1962. *The Human Skeleton in Forensic Medicine*. Springfield (Illinois): C. C. Thomas.
- Langendorfer S. J., Robertson M. A. 2002. Individual pathways in the development of forceful throwing. *Research Quarterly for Exercise and Sport* 73:245–256.
- Larson S. G. 2007a. The definition of humeral torsion: a comment on Rhodes (2006). *American Journal of Physical Anthropology* 133:819–820.
- Larson S. G. 2007b. Evolutionary transformation of the hominin shoulder. *Evolutionary Anthropology* 16:172–187.

- Larson S. G. 2009. Evolution of the hominin shoulder: early *Homo*. Pages 65–75 in *The First Humans: Origin and Early Evolution of the Genus Homo*, edited by F. E. Grine, J. G. Fleagle, and R. E. Leakey. New York: Springer.
- Larson S. G. 2012. Did Australopiths climb trees? *Science* 338:478–479.
- Larson S. G. 2015. Humeral torsion and throwing proficiency in early human evolution. *Journal of Human Evolution* 85:198–205.
- Larson S. G., Jungers W. L., Morwood M. J., Sutikna T., Jatmiko, Saptomo E. W., Due R. A., Djubiantono T. 2007. *Homo floresiensis* and the evolution of the hominin shoulder. *Journal of Human Evolution* 53: 718–731.
- Lassek W. D., Gaulin S. J. C. 2009. Costs and benefits of fat-free muscle mass in men: relationship to mating success, dietary requirements, and native immunity. *Evolution and Human Behavior* 30:322–328.
- Law D. J., Pellegrino J. W., Mitchell S. R., Fischer S. C., McDonald T. P., Hunt E. B. 1993. Perceptual and cognitive factors governing performance in comparative arrival-time judgments. *Journal of Experimental Psychology: Human Perception and Performance* 19:1183–1199.
- Lohman T. G., Roche A. F., Marotell R. 1988. *Anthropometric Standardization Reference Manual*. Champaign (Illinois): Human Kinetics Books.
- Lombardo M. P. 2012. On the evolution of sport. *Evolutionary Psychology* 10:1–28.
- Lombardo M. P., Deaner R. O. 2016. Anatomical adaptations for fighting. Pages 1–4 in *Encyclopedia of Evolutionary Psychological Science*, edited by T. K. Shackelford and V. Weekes-Shackelford. New York: Springer.
- Lombardo M. P., Deaner R. O. 2018. Born to throw: the ecological causes that shaped the evolution of throwing behavior in humans. *Quarterly Review of Biology* 93:1–16.
- Lordkipanidze D., Jashashvili T., Vekua A., Ponce de León M. S., Zollikofer C. P. E., Rightmire G. P., Pontzer H., Ferring R., Oms O., Tappen M., Bukhsianidze M., Agusti J., Kahlke R., Kiladze G., Martínez-Navarro B., Moushkelishvili A., Nioradze M., Rook L. 2007. Postcranial evidence from early *Homo* from Dmanisi, Georgia. *Nature* 449:305–310.
- Lorson K. M., Stoden D. F., Langendorfer S. J., Goodway J. D. 2013. Age and gender differences in adolescent and adult overarm throwing. *Research Quarterly for Exercise and Sport* 84:239–244.
- Lovejoy C. O., Simpson S. W., White T. D., Asfaw B., Suwa G. 2009. Careful climbing in the Miocene: the forelimbs of *Ardipithecus ramidus* and humans are primitive. *Science* 326:70–70e8.
- Low B. S. 1989. Cross-cultural patterns in the training of children: an evolutionary perspective. *Journal of Comparative Psychology* 103:311–319.
- Low B. S. 2000. *Why Sex Matters: A Darwinian Look at Human Behavior*. Princeton (New Jersey): Princeton University Press.
- Loy J. W., Hesketh G. L. 1995. Competitive play on the plains: an analysis of games and warfare among Native American warrior societies, 1800–1850. Pages 73–105 in *The Future of Play Theory: A Multidisciplinary Inquiry Into the Contributions of Brian Sutton-Smith*, edited by A. D. Pellegrini. Albany (New York): State University of New York Press.
- Lundh T., Udagawa J., Hänel S.-E., Otani H. 2011. Cross- and triple-ratios of human body parts during development. *Anatomical Record* 294:1360–1369.
- Macaluso P. J. Jr. 2011. Sex discrimination from the glenoid cavity in black South Africans: morphometric analysis of digital photographs. *International Journal of Legal Medicine* 125:773–778.
- Maki J. M. 2013. The biomechanics of spear throwing: an analysis of the effects of anatomical variation on throwing performance, with implications for the fossil record. PhD diss., Washington University.
- Malina R. M., Zavaleta A. N. 1976. Androgyny of physique in female track and field athletes. *Annals of Human Biology* 3:441–446.
- Marlowe F. M. 2007. Hunting and gathering: the human sexual division of foraging labor. *Cross-Cultural Research* 41:170–195.
- Marques-Bruna P., Grimshaw P. N. 1997. 3-dimensional kinematics of overarm throwing action of children age 15 to 30 months. *Perceptual and Motor Skills* 84: 1267–1283.
- Masharawi Y., Dar G., Peleg S., Steinberg N., Medlej B., May H., Abbas J., Hershkovitz I. 2010. A morphological adaptation of the thoracic and lumbar vertebrae to lumbar hyperlordosis in young and adult females. *European Spine Journal* 19:768–773.
- Matsumura N., Ogawa K., Kobayashi S., Oki S., Watanabe A., Ikegami H., Toyama Y. 2014. Morphologic features of humeral head and glenoid version in the normal glenohumeral joint. *Journal of Shoulder and Elbow Surgery* 23:1724–1730.
- Mayhew J. L., Salm P. 1990. Gender differences in anaerobic power tests. *European Journal of Applied Physiology and Occupational Physiology* 60:133–138.
- McKenzie T. L., Sallis J. F., Broyles S. L., Zive M. M., Nader P. R., Berry C. C., Brennan J. J. 2002. Childhood movement skills: predictors of physical activity in Anglo American and Mexican American adolescents? *Research Quarterly for Exercise and Sport* 73:238–244.
- Merrill A., Guzman K., Miller S. L. 2009. Gender differences in glenoid anatomy: an anatomic study. *Surgical and Radiologic Anatomy* 31:183–189.
- Meyer C., Nicklisch N., Held P., Fritsch B., Alt K. W. 2011. Tracing patterns of activity in the human skeleton: an overview of methods, problems, and limits

- of interpretation. *HOMO—Journal of Comparative Human Biology* 62:202–217.
- Miller A. E. J., MacDougall J. D., Tarnopolsky M. A., Sale D. G. 1993. Gender differences in strength and muscle fiber characteristics. *European Journal of Applied Physiology and Occupational Physiology* 66:254–262.
- Montoye H. J., Epstein F. H., Kjelsberg M. O. 1965. The measurement of body fatness: a study in a total community. *American Journal of Clinical Nutrition* 16: 417–427.
- Morrow J. R. Jr., Dische J. G., Ward P. E., Donovan T. J. III, Katch F. I., Katch V. L., Weltman A. L., Tellez T. 1982. Anthropometric, strength, and performance characteristics of American world class throwers. *Journal of Sports Medicine and Physical Fitness* 1:73–79.
- Murdock G. P., Provost C. 1973. Factors in the division of labor by sex: a cross-cultural analysis. *Ethnology* 12:203–225.
- Murphy A. M. C. 1994. Sex determination of prehistoric New Zealand clavicles. *New Zealand Journal of Archaeology* 16:85–91.
- Murphy A. M. C. 2002. Articular surfaces of the pectoral girdle: sex assessment of prehistoric New Zealand Polynesian skeletal remains. *Forensic Science International* 125:134–136.
- Nagarchi K., Pillai T. J., Saheb S. H., Brekeit K., Alharbi M. 2014. Morphometry of clavicle. *Journal of Pharmaceutical Sciences and Research* 6:112–114.
- National Center for Health Statistics. 16 June 2017. Clinical growth charts. Atlanta (Georgia): Centers for Disease Control and Prevention. http://www.cdc.gov/growthcharts/clinical_charts.htm.
- Nelson J. K., Thomas J. R., Nelson K. R., Abraham P. C. 1986. Gender differences in children's throwing performance: biology and environment. *Research Quarterly for Exercise and Sport* 57:280–287.
- Nordander C., Ohlsson K., Åkesson I., Arvidsson I., Balogh I., Hansson G.-Å., Strömberg U., Rittner R., Skerfving S. 2009. Risk of musculoskeletal disorders among females and males in repetitive/constrained work. *Ergonomics* 52:1226–1239.
- Osbahr D. C., Cannon D. L., Speer K. P. 2002. Retroversion of the humerus in the throwing shoulder of college baseball pitchers. *American Journal of Sports Medicine* 30:347–353.
- Oswalt W. H. 1976. *An Anthropological Analysis of Food-Getting Technology*. New York: John Wiley and Sons.
- Otterbein K. F. 2004. *How War Began*. College Station (Texas): Texas A&M University Press.
- Özer I., Katayama K., Sahgir M., Güleç E. 2006. Sex determination using the scapula in medieval skeletons from East Anatolia. *Collegium Anthropologicum* 30:415–419.
- Papaioannou V. A., Kranioti E. F., Joveniaux P., Nathana D., Michalodimitrakis M. 2012. Sexual dimorphism of the scapula and the clavicle in a contemporary Greek population: applications in forensic identification. *Forensic Science International* 217:231.e1–231.e7.
- Peters M. 1997. Gender differences in intercepting a moving target by using a throw or button press. *Journal of Motor Behavior* 29:290–296.
- Peters M., Mackenzie K., Bryden P. 2002. Finger length and distal finger extent patterns in humans. *American Journal of Physical Anthropology* 117:209–217.
- Petranek L. J., Barton G. V. 2011. The overarm-throwing pattern among U-14 ASA female softball players: a comparative study of gender, culture, and experience. *Research Quarterly for Exercise and Sport* 82:220–228.
- Pieper H.-G. 1998. Humeral torsion in the throwing arm of handball players. *American Journal of Sports Medicine* 26:247–253.
- Plomin R., DeFries J. C., McClearn G. E., McGuffin P. 2008. *Behavioral Genetics*. Fifth Edition. New York: Worth Publishers.
- Prescher A., Klumpen T. 1995. Does the area of the glenoid cavity of the scapula show sexual dimorphism? *Journal of Anatomy* 186:223–226.
- President's Commission on Campus Unrest. 1988. *The Kent State Tragedy: Special Report*. Salem (New Hampshire): Ayer Company Publishers.
- Puts D. A. 2010. Beauty and the beast: mechanisms of sexual selection in humans. *Evolution and Human Behavior* 31:157–175.
- Pyne D. B., Duthie G. M., Saunders P. U., Petersen C. A., Portus M. R. 2006. Anthropometric and strength correlates of fast bowling speed in junior and senior cricketers. *Journal of Strength and Conditioning Research* 20:620–626.
- Rhodes J. A. 2006. Adaptations to humeral torsion in medieval Britain. *American Journal of Physical Anthropology* 130:160–166.
- Rhodes J. A., Churchill S. E. 2009. Throwing in the Middle and Upper Paleolithic: inferences from an analysis of humeral retroversion. *Journal of Human Evolution* 56:1–10.
- Rhodes J. A., Knüsel C. J. 2005. Activity-related skeletal change in medieval humeri: cross-sectional and architectural alterations. *American Journal of Physical Anthropology* 128:536–546.
- Rippee N. E., Pangrazi R. P., Corbin C. B., Borsdorf L., Petersen G., Pangrazi D. L. 1990. Throwing profiles of first and fourth grade boys and girls. *Physical Educator* 47:180–185.
- Roach N. T., Lieberman D. E. 2014. Upper body contributions to power generation during rapid, overhand throwing in humans. *Journal of Experimental Biology* 217:2139–2149.
- Roach N. T., Richmond B. G. 2015. Clavicle length, throwing performance and the reconstruction of the *Homo erectus* shoulder. *Journal of Human Evolution* 80:107–113.

- Roach N. T., Lieberman D. E., Gill T. J., Palmer W. E., Gill T. J. 2012. The effect of humeral torsion on rotational range of motion in the shoulder and throwing performance. *Journal of Anatomy* 220:293–301.
- Roach N. T., Venkadesan M., Rainbow M. J., Lieberman D. E. 2013. Elastic energy storage in the shoulder and the evolution of high-speed throwing in *Homo*. *Nature* 498:483–486.
- Rowe K. 20 November 2012. How We Practice: Throwing Like a Girl. <https://ultiworld.com/2012/11/20/how-we-practice-throwing-like-a-girl/>.
- Ruff C. B. 1991. Climate and body shape in hominid evolution. *Journal of Human Evolution* 21:81–105.
- Ruff C. B. 1994. Morphological adaptation to climate in modern and fossil hominids. *Yearbook of Physical Anthropology* 37:65–107.
- Ruff C. B. 2003. Growth in bone strength, body size, and muscle size in a juvenile longitudinal sample. *Bone* 33:317–329.
- Sakurai S., Miyashita M. 1983. Developmental aspects of overarm throwing related to age and sex. *Human Movement Science* 2:67–76.
- Sallis R. E., Jones K., Sunshine S., Smith G., Simon L. 2001. Comparing sports injuries in men and women. *International Journal of Sports Medicine* 22:420–423.
- Sanders G., Sinclair K. 2011. Sex differences in accuracy and precision when judging time to arrival: data from two internet studies. *Archives of Sexual Behavior* 40:1189–1198.
- Schiff W., Oldak R. 1990. Accuracy of judging time to arrival: effects of modality, trajectory, and gender. *Journal of Experimental Psychology: Human Perception and Performance* 16:303–316.
- Shea J. J. 2009. The impact of projectile weaponry on Late Pleistocene hominin evolution. Pages 189–199 in *The Evolution of Hominin Diets: Integrating Approaches to the Study of Palaeolithic Subsistence*, edited by J.-J. Hublin and M. P. Richards. Dordrecht (The Netherlands): Springer.
- Shobha S. C. M., Jatti V. B., Murthy C. R. V. 2014. Identification of sex of human clavicles from North Karnataka Zone. *Anthropologist* 17:917–920.
- Shulman J. L., Bowen W. G. 2001. *The Game of Life: College Sports and Educational Values*. Princeton (New Jersey): Princeton University Press.
- Shumaker R. W., Walkup K. R., Beck B. B. 2011. *Animal Tool Behavior: The Use and Manufacture of Tools By Animals*. Baltimore (Maryland): Johns Hopkins University Press.
- Singh S., Gangrade K. C. 1968. The sexing of adult clavicles—demarking points for Varanasi zone. *Journal of Anatomical Society of India* 17:89–100.
- Sipes R. G. 1973. War, sports, and aggression: an empirical test of two rival theories. *American Anthropologist* 75:64–86.
- Smith E. A. 2004. Why do good hunters have higher reproductive success? *Human Nature* 15:343–364.
- Stern J. T. Jr. 2000. Climbing to the top: a personal memoir of *Australopithecus afarensis*. *Evolutionary Anthropology* 9:113–133.
- Stewart T. D. 1979. *Essentials of Forensic Anthropology, Especially as Developed in the United States*. Springfield (Illinois): Thomas.
- Steyn M., İşcan M. Y. 1997. Sex determination from the femur and tibia in South African whites. *Forensic Science International* 90:111–119.
- Stodden D. F., Langendorfer S. J., Fleisig G. S., Andrews J. R. 2006a. Kinematic constraints associated with the acquisition of overarm throwing. Part I: step and trunk actions. *Research Quarterly for Exercise and Sport* 77:417–427.
- Stodden D. F., Langendorfer S. J., Fleisig G. S., Andrews J. R. 2006b. Kinematic constraints associated with the acquisition of overarm throwing. Part II: upper extremity actions. *Research Quarterly for Exercise and Sport* 77:428–436.
- Stoudt H. W., Damon A., McFarland R. A. 1970. Skinfolts, body girths, biacromial diameter, and selected anthropometric indices of adults. *National Center for Health Statistics Vital and Health Statistics* 11:1–69.
- Sugiyama Y., Koman J. 1979. Tool-using and -making behavior in wild chimpanzees at Bossou, Guinea. *Primates* 20:513–524.
- Tanner J. M. 1978. *Foetus Into Man: Physical Growth From Conception to Maturity*. Cambridge (Massachusetts): Harvard University Press.
- Terry R. J. 1932. The clavicle of the American Negro. *American Journal of Physical Anthropology* 16:351–379.
- Testart A. 1986. *Essai sur les fondements de la division sexuelle du travail chez les chasseurs-cueilleurs*. Paris (France): Editions de l'Ecole des Hautes Etudes en Sciences Sociales.
- Thomas J. R., French K. E. 1985. Gender differences across age in motor performance: a meta-analysis. *Psychological Bulletin* 98:260–282.
- Thomas J. R., Marzke M. W. 1992. The development of gender differences in throwing: is human evolution a factor? Pages 60–76 in *Enhancing Human Performance in Sport: New Concepts and Development*, edited by R. W. Christina and H. M. Eckert. Champaign (Illinois): Human Kinetics Publishers.
- Thomas J. R., Thomas K. T. 1988. Development of gender differences in physical activity. *Quest* 40: 219–229.
- Thomas J. R., Michael D., Gallagher J. D. 1994. Effects of training on gender differences in overhand throwing: a brief quantitative literature analysis. *Research Quarterly for Exercise and Sport* 65:67–71.
- Thomas J. R., Alderson J. A., Thomas K. T., Campbell A. C., Elliott B. C. 2010. Developmental gender differences for overhand throwing in aboriginal Australian children. *Research Quarterly for Exercise and Sport* 81:432–441.

- Turney-High H. H. 1971. *Primitive War: Its Practice and Concepts*. Columbia (South Carolina): University of South Carolina Press.
- van Creveld M. 1989. *Technology and War: From 2000 B.C. to the Present*. New York: Free Press.
- van den Tillaar R., Ettema G. 2004. Effect of body size and gender in overarm throwing performance. *European Journal of Applied Physiology* 91:413–418.
- Vandaele B., Cools W., de Decker S., de Martelaer K. 2011. Mastery of fundamental movement skills among 6-year-old Flemish pre-school children. *European Physical Education Review* 17:3–17.
- Veldman S. L. C., Palmer K. K., Okely A. D., Robinson L. E. 2017. Promoting ball skills in preschool-age girls. *Journal of Science and Medicine in Sport* 20:50–54.
- Venetsanou F., Kambas A. 2016. Motor proficiency in young children: a closer look at potential gender differences. *SAGE Open* January–March 2016:1–10.
- Villotte S., Knüsel C. J. 2014. “I sing of arms and of a man . . .”: medial epicondylitis and the sexual division of labour in prehistoric Europe. *Journal of Archaeological Science* 43:168–174.
- Waidhofer M., Kirchengast S. 2015. Sexual dimorphism in directional asymmetry of the upper limb bones among Khoe-San skeletons. *HOMO—Journal of Comparative Human Biology* 66:508–519.
- Walker A., Leakey R. E. 1993. *The Nariokotome Homo erectus Skeleton*. Cambridge (Massachusetts): Harvard University Press.
- Watson N. V., Kimura D. 1988. Throwing accuracy: lateralization, sex differences, and spatial ability. *Canadian Psychology* 29:830.
- Watson N. V., Kimura D. 1991. Nontrivial sex differences in throwing and intercepting: relation to psychometrically-defined spatial functions. *Personality and Individual Differences* 12:375–385.
- West-Eberhard M. J. 2003. *Developmental Plasticity and Evolution*. Oxford (United Kingdom): Oxford University Press.
- Whitcome K. K., Shapiro L. J., Lieberman D. E. 2007. Fetal load and the evolution of lumbar lordosis in bipedal hominins. *Nature* 450:1075–1078.
- Whiteley R. J. 2007. Baseball throwing mechanics as they relate to pathology and performance—a review. *Journal of Sports Science and Medicine* 6:1–20.
- Whiteley R. J., Ginn K. A., Nicholson L. L., Adams R. D. 2009. Sports participation and humeral torsion. *Journal of Orthopaedic & Sports Physical Therapy* 39:256–263.
- Wickstrom R. L. 1977. *Fundamental Motor Patterns*. Philadelphia (Pennsylvania): Lea & Febiger.
- Wild M. R. 1938. The behavior pattern of throwing and some observations concerning its course of development in children. *Research Quarterly for Exercise and Sport* 9:20–24.
- Williams G. C. 1985. A defense of reductionism in evolutionary biology. *Oxford Surveys in Evolutionary Biology* 2:1–27.
- Williams H. G., Monsma E. V. 2007. Assessment of gross motor development. Pages 397–434 in *Psychological Assessment of Preschool Children*, Fourth Edition, edited by B. A. Bracken and R. J. Nagle. Hillsdale (New Jersey): Lawrence Erlbaum Associates.
- Williams K., Haywood K. M., Painter M. A. 1996. Environmental versus biological influences on gender differences in the overarm throw for force: dominant and nondominant arm throws. *Women in Sport and Physical Activity Journal* 5:29–48.
- Wilson E. O. 1998. *Consilience: The Unity of Knowledge*. New York: Knopf.
- Wilson M., Daly M. 1985. Competitiveness, risk taking, and violence: the young male syndrome. *Ethology and Sociobiology* 6:59–73.
- Wood W., Eagly A. H. 2002. A cross-cultural analysis of the behavior of women and men: implications for the origins of sex differences. *Psychological Bulletin* 128:699–727.
- Wyland D. J., Pill S. G., Shanley E., Clark J. C., Hawkins R. J., Noonan T. J., Kissenberth M. J., Thigpen C. A. 2012. Bony adaptation of the proximal humerus and glenoid correlate within the throwing shoulder of professional baseball pitchers. *American Journal of Sports Medicine* 40:1858–1862.
- Young I. M. 1980. Throwing like a girl: a phenomenology of feminine body comportment motility and spatiality. *Human Studies* 3:137–156.
- Young I. M. 1998. “Throwing like a girl”: twenty years later. Pages 286–290 in *Body and Flesh: A Philosophical Reader*, edited by D. Welton. Malden (Massachusetts): Blackwell Publishing.
- Young N. M., Capellini T. D., Roach N. T., Alemseged Z. 2015. Fossil hominin shoulders support an African ape-like last common ancestor of humans and chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America* 112:11829–11834.
- Young R. W. 2003. Evolution of the human hand: the role of throwing and clubbing. *Journal of Anatomy* 202:165–174.
- Young R. W. 2009. The ontogeny of throwing and striking. *Human Ontogenetics* 3:19–31.
- Zerjal T., Xue Y., Bertorelle G., et al. 2003. The genetic legacy of the Mongols. *American Journal of Human Genetics* 72:717–721.
- Zhu Q., Dapena J., Bingham G. P. 2009. Learning to throw to maximum distances: do changes in release angle and speed reflect affordances for throwing? *Human Movement Science* 28:708–725.

Handling Editor: Daniel E. Dykhuizen

Implications of gender-affirming endocrine care for sports participation

Ethan Moreland, Ada S. Cheung^{ID}, Danielle Hiam, Brendan J. Nolan, Shanie Landen, Magsue Jacques, Nir Eynon and Patrice Jones^{ID}

Ther Adv Endocrinol Metab

2023, Vol. 14: 1–17

DOI: 10.1177/
20420188231178373

© The Author(s), 2023.
Article reuse guidelines:
sagepub.com/journals-
permissions

Abstract: Many transgender (trans) individuals utilize gender-affirming hormone therapy (GAHT) to promote changes in secondary sex characteristics to affirm their gender. Participation rates of trans people in sport are exceedingly low, yet given high rates of depression and increased cardiovascular risk, the potential benefits of sports participation are great. In this review, we provide an overview of the evidence surrounding the effects of GAHT on multiple performance-related phenotypes, as well as current limitations. Whilst data is clear that there are differences between males and females, there is a lack of quality evidence assessing the impact of GAHT on athletic performance. Twelve months of GAHT leads to testosterone concentrations that align with reference ranges of the affirmed gender. Feminizing GAHT in trans women increases fat mass and decreases lean mass, with opposite effects observed in trans men with masculinizing GAHT. In trans men, an increase in muscle strength and athletic performance is observed. In trans women, muscle strength is shown to decrease or not change following 12 months of GAHT. Haemoglobin, a measure of oxygen transport, changes to that of the affirmed gender within 6 months of GAHT, with very limited data to suggest possible reductions in maximal oxygen uptake as a result of feminizing GAHT. Current limitations of this field include a lack of long-term studies, adequate group comparisons and adjustment for confounding factors (e.g. height and lean body mass), and small sample sizes. There also remains limited data on endurance, cardiac or respiratory function, with further longitudinal studies on GAHT needed to address current limitations and provide more robust data to inform inclusive and fair sporting programmes, policies and guidelines.

Keywords: gender-affirming hormone therapy, muscle, sports, transgender

Received: 31 August 2022; revised manuscript accepted: 9 May 2023.

Introduction

Transgender (trans) is an umbrella term for individuals with a gender identity different from that presumed at birth and includes people with binary (man/woman) and non-binary identities. Gender affirmation, including social, legal and/or medical affirmation, plays an important role in improving the mental health of trans people.^{1–3} Medical affirmation involving gender-affirming hormone therapy (GAHT) results in clear changes in circulating sex hormone concentrations, which allows for the development of secondary sex characteristics more aligned to one's affirmed gender.^{4,5} Feminizing GAHT typically involves oestradiol

and anti-androgen administration, which induces breast formation, reduces male pattern hair growth, and alters fat and muscle distribution.^{6,7} Masculinizing GAHT involves testosterone therapy and stimulates increases in facial and body hair, changes in fat and lean mass distribution and longer-term effects such as deepening of the voice.^{6,7} While the effects of GAHT on secondary sex characteristics are well-established, the broader physiological effects of these hormone therapies in various tissues are less clear. In this respect, there is current interest in the impacts of GAHT on skeletal muscle and on measures relevant to participation, performance and safety in sports.

Correspondence to:

Patrice Jones
Institute for Health and
Sport (IHeS), Victoria
University, Ballarat Road,
Footscray, VIC 3011,
Australia

Patrice.Jones@vu.edu.au

Ethan Moreland
Shanie Landen
Magsue Jacques
Nir Eynon
Institute for Health and
Sport (IHeS), Victoria
University, Footscray, VIC,
Australia

Ada S. Cheung
Brendan J. Nolan
Trans Health Research
Group, Department
of Medicine (Austin
Health), The University of
Melbourne, Heidelberg,
VIC, Australia

Department of
Endocrinology, Austin
Health, Heidelberg, VIC,
Australia

Danielle Hiam
Institute for Physical
Activity and Nutrition
(IPAN), School of Exercise
and Nutrition Sciences,
Deakin University,
Burwood, VIC, Australia

*Nir Eynon is also affiliated
to Australian Regenerative
Medicine Institute, Monash
University, Clayton,
VIC, Australia; Shanie
Landen is also affiliated
to Hudson Institute for
Medical Research, Monash
University, Clayton, VIC,
Australia



There is ongoing debate around whether trans individuals should be included in grass-roots and elite sport competitions. The premise of competitive sport, whether at the grass-roots and elite level, is based on subjective fairness, with the goal of inclusion for everyone, in the absence of discrimination.⁸ Many sports are separated into male and female categories due to evident sex differences in performance, and concerns for safety of female athletes in contact sports. On average, males are stronger than females and possess larger body sizes, muscle mass, and whole-body oxygen consumption.^{9,10} These differences contribute to differences in male and female performance, that range from 5 to 17%, depending on the sport and competition level.¹¹ Due to limited evidence and research in trans health, it remains unclear whether there is an athletic advantage or disadvantage for trans men or trans women after undergoing GAHT in different sport disciplines. Participation in sport is a basic right for all people from any background. It is therefore important to understand whether GAHT confers any advantage or disadvantage in certain sport settings, to inform sports policies and programmes.¹² Here, we review the current knowledge and limitations around the effects of GAHT on measures relevant to sports.

Terminology

In this review, *male* and *female* are sex terms referring to someone's sex presumed at birth. *Woman/women*, *man/men*, *non-binary* and *gender-diverse* are gender terms which describe a person's gender identity. Someone who has a gender identity that aligns with their sex at birth may be referred to as *cisgender* (cis). *Transgender* (trans) describes people who have a gender identity different from the sex they were assigned at birth. This review focuses on studies of transgender men and women using standard doses of GAHT to achieve sex steroid concentrations in the range of the affirmed gender and does not cover people who may be using low doses of GAHT, given the lack of relevant studies.

Brief view of transgender participation in sport

Trans athletes throughout history have undergone discrimination and exclusion in their given sports, despite trans people representing a small

percentage of the general population (0.1–2.0%¹³) and sports competitors. As examples, trans women Renée Richards and Fallon Fox were banned from competing in their respective sports (1976: US Open Tennis Woman's Championship, 2013: mixed martial arts) due to perceived performance advantages.^{14,15} The International Olympic Committee have allowed trans athletes to compete since 2003, provided certain criteria are met. In that time, despite over 70,000 individuals competing in the Olympic Games, only a handful of trans athletes have competed, and none have medalled; however, these athletes' faced discrimination and public debate despite abiding by regulations and rules. More recently, in 2021, weightlifter Laurel Hubbard became the first openly trans woman to compete in an Olympics Game and was subjected to online abuse and media hounding.¹⁶ Trans men athletes have typically been the subject of less public debate; however, Chris Mosier, known for being the first trans man to compete in a men's Olympic Trials in 2020, has spoken openly about the challenges faced to be eligible for these competitions.^{17,18}

The increased visibility of trans people, in general, has fuelled ongoing public debates and policy actions around the inclusion of trans people in sport. Previous sporting guidelines have been largely focused on the serum testosterone concentrations of competing trans women athletes. In 2015, the International Olympic Committee (IOC) restricted many trans women athletes from competing in the women's category by introducing a testosterone threshold of below 10nmol/L for a minimum of 12 months prior to competition.¹⁹ Trans men were permitted to compete without restriction following 1 year of GAHT. However, in November 2021, the IOC released a new framework for participation, which encouraged moving away from testosterone-based restrictions and a 'one-size-fits-all' approach.⁸ The framework stated that sporting bodies should not assume that trans women have an inherent advantage over cis women, nor should they have to reduce their testosterone levels to compete. The IOC also concluded that individual sporting bodies should hold the responsibility of establishing guidelines for trans inclusion for their given sport, with guidelines based on robust and peer-reviewed evidence.⁸ While this approach makes logical sense, the issue is that it is practically impossible given the current lack of evidence in this field.

During the writing of this narrative review, FINA, the governing body of international competitive swimming, announced a 2022 policy stating that trans athletes who transitioned after puberty are ineligible to compete in elite international swimming competitions.²⁰ FINA stated that the decision was made after consulting scientists and policymakers, yet there is currently no definitive research in trans swimmers which implies any performance enhancement in trans women. Sporting organizations currently appear pressured to make evidence-based policy decisions around trans participation, with the prevailing problem being there is very little quality research. Given that sport is an important fabric of society in many countries, there are also current concerns about how decisions being made at the elite sporting level will impact community sport guidelines and broader motivations of trans people to engage in sports.

At the community level, many sporting guidelines aimed at boosting the participation of trans people in sport have been recently introduced. For example, in 2020 in Australia, several governing bodies, including that of Tennis, Rugby, Australian Rules Football, Hockey, Netball, Water Polo, Touch Football and university sports released guidelines for trans participation.²¹ However, despite the recent development of trans-specific guidelines, participation in sports is exceedingly low.²² Recent studies reveal trans individuals engage less in sports and physical activity, when compared to cis individuals,²³ and are deterred from competing in single-sex sports.^{24,25} A 2020 survey of Spanish trans individuals ($n=212$) identified gender disclosure as a key issue in sports participation, with 14.5% stopping physical activity after gender disclosure,²⁴ while it was reported 52% of boys and 44% of girls participated in sport,²⁶ compared to 12% of trans youth (14% for trans boys and 12% for trans girls) participate in sports teams.²² This is a concern when considering that trans people are at higher risk of poor physical and mental health,^{27,28} and would especially benefit from regular engagement in sports and exercise. Given many organizational barriers to sports participation stem from the unclear effects of GAHT used by trans individuals, further research and critical review in this field is needed to inform updated policies, guidelines and sporting programmes which are evidence-based, fair and more inclusive and beneficial for all.

Performance differences between sexes

Sex differences in performance between males and females are well reported and should be highlighted before discussions of differences in performance between cisgender individuals and trans individuals undergoing GAHT. A systematic review by Cheuvront *et al.* concluded that males held a considerable aerobic performance advantage over females, ranging from 8.4% (marathon distance) to 14.1% (5000m).²⁹ Performance advantages of 7.3–11.9% were also reported for males in sprint-based events of 100–1500m.²⁹ A 2007 review and comparison of the top male and female athletes in Olympic and World Championship events from 1952 to 2006 concluded that performance differences between sexes have widened since the 1990s in running, swimming and speed skating events.³⁰ A study looking at swimming, running and jumping events reported the sex differences from the onset of male puberty to early adulthood. For track and field athletics, the study reported pre-pubertal sex differences from 3.0% to 10.1% across events. For jump performance, the difference between sexes pre-pubertal was ~5.8%, which increased to a 19.4% advantage for males after the onset of puberty.³¹ Sex differences are also seen in marathon events, with a 2018 study finding that females achieved their best marathon race times 5 years earlier than males, but males outperformed females in their best race times.³² An earlier study comparing marathon performance in matched males and females found significant differences in running economy, with females having poorer running economy and training volume over males.³³ A 1998 review of the world's best times in the 1500m and marathon events from 1980 to 1996 states that sex difference in these events has plateaued at 11.2%.³⁴

Performance differences between sexes are underpinned by sex variation in several measures, such as body composition, strength and aerobic capacity. In terms of strength, there is a large disparity in strength between adult males and females, which is a major reason for single-sex sporting competitions. Males are evidently stronger than females, with a comparative study concluding that females have 37–68% of the muscle strength of that of males, with the biggest disparity being in the upper body.³⁵ As such, it would not be meaningful for males and females to compete together, particularly in impact sports. There are further sex differences in aerobic capacity,

measured by maximal oxygen uptake ($\text{VO}_{2\text{max}}$), with females having been found to have 70–75% of the capacity of males.³⁶ This seems to be a significant factor in the differences in race times outlined above, particularly for middle and long-distance events.³⁷ In terms of strength, it is well known that males have a greater capacity for maximal strength tasks such as weightlifting.³⁸ One study performed two correlational observations, during which they found that in the first observation of national-level USA weightlifters, males ($n=39$) displayed greater strength relative to bodyweight in the squat (22%), snatch (27%) and clean (28%) movements compared to females ($n=26$).³⁸ Males and females also differ in body composition,³⁹ with one prospective study showing that males ($n=94$) possessed 20% greater lean mass/weight ratio than females ($n=114$), while having 26% less total fat mass when expressed as a percentage of body weight.³⁹ Fracture rates also differ between males and females,^{40,41} which may be attributed to the higher bone mineral density (BMD) and content (BMC) in males compared to females,⁴⁰ and have implications for performance and safety in weight-bearing and endurance sports. One study of height and weight-matched groups of 18-year-old males ($n=36$) and females ($n=36$) found significantly greater BMC for males at the femoral neck (11%), femoral trochanter (27%), femoral shaft (10%), total proximal femur (16%) and proximal tibia (13%).⁴² This study also showed significantly greater BMD in femoral neck (6%), femoral trochanter (6%), femoral shaft (7%) and proximal tibia (7%) in males compared to weight and height-matched females.⁴²

Effects of GAHT on serum sex hormone concentrations

Gender-affirming hormone therapies have clear effects on circulating sex hormone concentrations, with testosterone the focal hormone in sports, given its clear association with muscle strength and performance in cisgender males.⁴³ Testosterone levels in healthy cisgender males (8.8–30.9 nmol/L) are four to fivefold greater than the upper end of the typical cisgender female range (0.4–2.0 nmol/L) and contribute significantly to biological sex differences.⁴⁴ In males, testosterone influences athletic performance by promoting the development and maintenance of muscle mass and reductions in body fat.⁴⁵ In females, this relationship is less clear, with mixed

results in the literature.⁴⁶ While evidence suggests testosterone is not a perfect proxy for performance, many sporting eligibility guidelines continue to focus on testosterone concentrations in both men and women. Despite limitations in using testosterone thresholds in sports, there is good evidence supporting the efficacy of feminizing therapies in lowering testosterone concentrations to the typical female reference range in trans women, and masculinizing hormone therapies in increasing serum testosterone concentrations to the typical male reference range in trans men.⁴⁷

Feminizing gender-affirming hormone therapies involve oestrogen administration, often in combination with anti-androgen agents such as cyproterone acetate (CPA), spironolactone or gonadotropin-releasing hormone analogues (e.g. leuprolide).⁴⁷ A 2020 systematic review found that anti-androgens CPA, medroxyprogesterone acetate (MPA) and leuprolide may be more effective than spironolactone at suppressing serum total testosterone concentrations when combined with oestrogen, but this is likely because of differences in the mechanism of action with spironolactone blocking peripheral androgen receptors rather than reducing testosterone production.⁴⁸ For example, a 2019 cross-sectional study of trans women using GAHT for a median of 1.5 years found those treated with oestradiol plus CPA (median dose 50 mg daily) had significantly lower median serum total testosterone concentration (0.8 nmol/L), compared to those on oestradiol plus spironolactone (2.0 nmol/L: median spironolactone dose 100 mg/day) or oestradiol alone (10.5 nmol/L: median oestradiol dose 6 mg/day). In this study, the median testosterone concentration in both CPA and spironolactone groups was within the typical female range (0.4–2.0 nmol/L).⁴⁹ Another study of trans women ($n=40$) receiving transdermal oestradiol (1–2 mg/day) with either 50 mg/day of CPA (oral) or 3.75 mg/month of leuprolide (intramuscular injection) reported that both therapies reduced serum total testosterone concentrations to within the typical female range following 12 months.⁵⁰ A 2021 study examined the effects of different doses of CPA paired with oestrogen in trans women ($n=882$, finding not only that 10 mg of CPA was as effective as higher doses (25, 50, 100 mg), but that circulating testosterone concentrations reduced to within female reference range within 3 months.⁵¹ Comparatively, a 2019 study found circulating testosterone concentrations of trans

women fell within the female reference range after 4 months of oestradiol treatment ($n=11$).⁵² In a sports context, these findings suggest that trans women may therefore meet set testosterone thresholds at different times in their treatment, depending on what treatment is being followed. However, it is important to highlight the limitations of sporting guidelines which enforce testosterone limits on trans women. Given that anti-androgens are commonly used in feminizing GAHT, it cannot be assumed that testosterone concentrations directly reflect testosterone activity in trans women. Anti-androgens work to both suppress testosterone production and inhibit testosterone activity at the receptor level. A trans woman using anti-androgens may therefore present with high testosterone concentrations during the early stages of GAHT, but the action of testosterone is being inhibited at the critical receptor level.

In comparison to feminizing GAHT, masculinizing GAHT involves testosterone administration, either intramuscular, transdermal, or subcutaneous and leads to serum testosterone concentrations increasing to within the typical cisgender male range (8.8–30.9 nmol/L) following 12 months.^{53–55} A 2015 study in trans men ($n=23$) found 12 months of masculinizing GAHT (1000 mg intramuscular testosterone undecanoate every 12 weeks) led to serum testosterone concentrations increasing from 0.9 to 21.8 nmol/L.⁵³ A later 2018 study ($n=50$ trans men)⁵⁴ reported that intramuscular injections of testosterone undecanoate (1000 mg at weeks 0 and 6, then every 12–16 weeks) or testosterone enanthate (250 mg every 3–4 weeks) also increased serum testosterone concentrations to within male ranges within 12 months, from 1.4 to 18.3 nmol/L and 1.9 to 18.0 nmol/L in the testosterone undecanoate and testosterone enanthate groups, respectively. A further randomized controlled trial found that testosterone gel (50 mg/day) was just as effective as injections of testosterone undecanoate (1000 mg at weeks 0 and 6, then every 12–16 weeks) and testosterone enanthate (100 mg every 10 days) at increasing total testosterone concentrations to within normal male range by week 30 of treatment.⁵⁵ While there are no current restrictions placed on men around testosterone concentrations, sport guidelines (e.g. FINA²⁰ and World Rugby⁵⁶) typically require trans men using testosterone (i.e. considered to be performance-enhancing substances) to obtain a Therapeutic Use Exemption (TUE) to allow

them to use to testosterone whilst competing. Without a TUE, trans men competitors risk committing an Anti-Doping Rule Violation and being suspended from their sport.

Effects of GAHT on body composition

Current evidence indicates that feminizing GAHT leads to increased fat mass and decreased lean mass in trans women (Figure 1), with inverse effects shown in trans men using masculinizing GAHT.^{57–63}

A recent systematic review found that feminizing hormone therapies decreased lean body mass by 0.8–5.4% following 12 months.⁵⁷ However, it is worth noting that is large variations in body composition changes are seen across individual studies. One study finds no change in body composition of 9.4% of trans women ($n=288$) following 24 months of GAHT.⁶⁴ A further prospective study in trans adolescent girls ($n=21$) found a reduction of 4.7% in lean mass over 10.6 months ($n_{\text{range}}=5\text{--}31$ months).⁶⁵ These changes differ by region, with a prospective study ($n=179$ trans women) finding the largest lean mass decreases occurred in the arm (–5 to –7%) and leg region (–3 to –4%), compared to gynoid regions (–2 to –3%) and android regions (–1% to +2%).⁵⁸ In terms of total body fat, increases of 9–32%^{58,66–68} have been observed following 12 months of feminizing GAHT, with the greatest increases seen in the leg region (+37 to +47%), and the smallest changes in the trunk region (+16 to +25%).^{58,65} A study in trans women ($n=11$) found increased regional body fat of 18, 27.4 and 27.2% in the trunk, legs and gynoid region, respectively, and an average 7.2% increase in total body fat of 7.2% following 6 months of feminizing GAHT.⁶⁹ In terms of muscle area, a retrospective cohort study of trans women ($n=17$) reported average reductions of 10% and 12% following 1 and 3 years of feminizing GAHT, respectively.⁷⁰ Further research reports an increase in body mass index (BMI) in trans women following GAHT, with three studies reporting increases of 9–10% over 1–7 years of GAHT use.^{59–61} A limitation of prior studies exploring body composition changes in trans women is that they lack adequate comparison groups of cis women, making it difficult to assess how body composition may differ across cis and trans women. Existing reference standards for body composition measures indicate that males and females differ by an average of

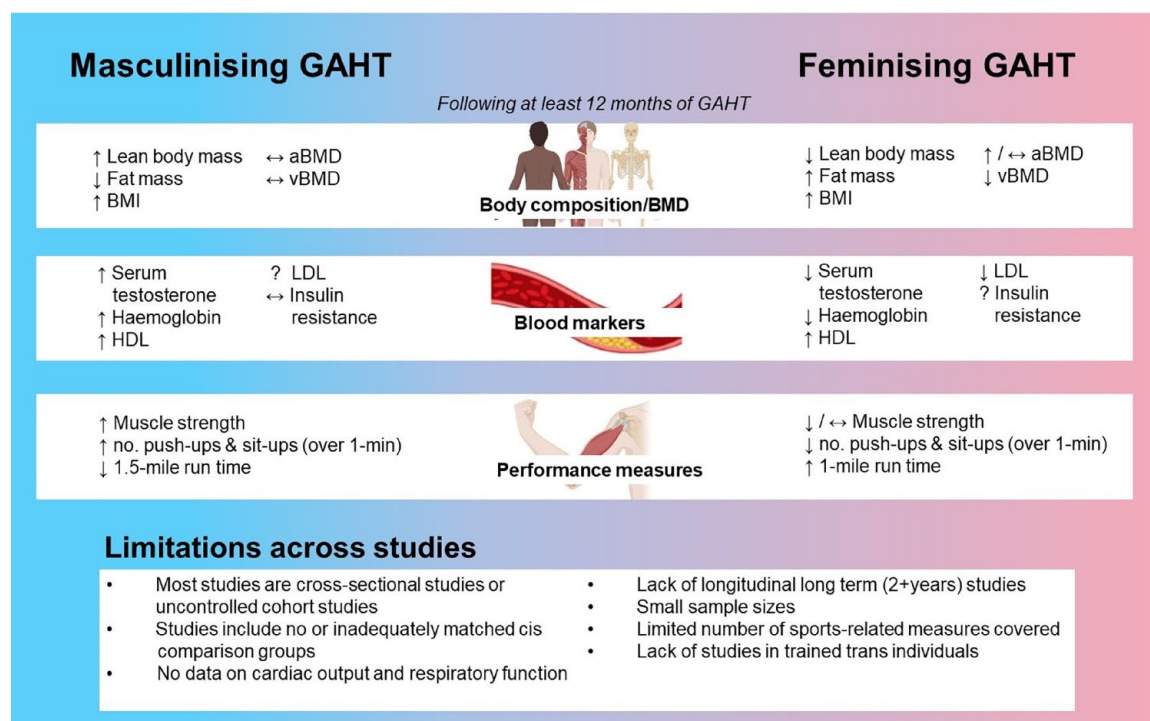


Figure 1. Summary and limitations of reported effects of GAHT on outcomes relevant to sports. aBMD, areal bone mineral density; BMI, body mass index; DMPs, differentially methylated positions; GAHT, gender-affirming hormone therapy; HDL, high-density lipoprotein; LDL, low-density lipoprotein; vBMD, volumetric bone mineral density.

+10-12% across all age groups in mean % lean body mass⁷¹ and -20% for fat mass.⁷² Acknowledging this, prior research indicates that trans women using GAHT, on average, increase their total fat mass to within the expected female ranges within 12 months. In addition, trans women also experience decreases in lean body mass, which, on average, remain higher than expected female values after 12 months. However, these comparisons do not account for differences in age, height and BMI, which need to be considered for accurate comparisons. These measures also determine body composition outcomes in trans women, with, for example, higher BMI at start of hormone therapy related to smaller increases in body fat and larger decreases in lean mass.⁵⁸

In comparison to feminizing GAHT, masculinizing GAHT leads to decreased fat mass and increased lean mass in trans men, with variation in degrees of change seen across individual studies. A prospective study examining the effects of GAHT over 12 months ($n=23$ trans men) found a 10.4% mean increase in lean mass and a 9.7% mean reduction in total body fat mass,⁵³ with

comparable findings seen in a 12 months prospective study ($n=53$ trans men, -13% fat mass, +11% lean mass),⁶⁶ but smaller changes seen in a retrospective study ($n=121$) (+3% lean mass, -3% fat mass) following 1.0 years of gonadotropin-releasing hormone analogues (GnRH) and 2.4 years GAHT.⁶⁸ In trans adolescents, one prospective study found an 8.6% increase in lean mass following a mean 11.6 months ($n_{\text{range}}=4-40$ months) of treatment in trans boys ($n=44$).⁶⁵ Another prospective study ($n=162$ trans men) reported similar findings, with a mean 10% decrease in total body fat and a 10% increase in lean mass over 12 months of GAHT. This study observed the largest body fat decreases in the leg (-14% to -19%) and gynoid regions (-12% to -16%), and the largest lean mass increases in the arm region (+18 to +21%).⁵⁸ Prior research also found that the BMI of trans men significantly increases following 6 months,^{60,62} and 12 months of GAHT⁵⁵ with increases of 9-10%, and 2% reported at these time points, respectively. Limited studies have gathered comparison data from cis men and women. A 2012 study in trans men ($n=50$) and cis women ($n=50$) matched for age, height and BMI found that, on average, trans

men had 30% less body fat and 9% more lean body mass compared to cis women. A limitation of this study was that it assessed trans men who had undergone gender affirmation surgery in addition to long-term GAHT and therefore, the separate effects of GAHT and gender confirmation surgery on body composition measures could not be determined. A more recent 2022 study compared body composition, among other measures, between trans men ($n=19$) and cis men ($n=19$) matched for age and BMI, and found trans men had 26.3% lower total lean mass and 18% higher body fat than cis men ($n=19$), following a minimum of 6 months of GAHT.⁶³ A retrospective study examined muscle area in those undergoing masculinizing GAHT ($n=19$), finding increases of 16% and 15% following 1 and 3 years of GAHT, respectively.⁷⁰ This study however did not have any comparison groups, so data could not be compared to cisgender individuals. These findings indicate that trans men likely, on average, possess lower muscle mass and higher body fat than cis men. This has potential safety implications, such as increased risk of concussion or injury, for trans men in high-contact sports. These risks are already acknowledged in many sporting guidelines. For example, World Rugby requires trans men to provide a confirmation of physical ability to play, which includes written acknowledgement and acceptance by the player of the associated risks of playing contact sports with males who are statistically likely to be stronger, faster and heavier than them, as well as written confirmation from a medical practitioner or qualified coach that the player is in a physical condition to play.⁵⁶

Effects of GAHT on muscle strength

A 2020 systematic review reported decreased or no change in muscle strength following at least 12 months of GAHT, across five longitudinal studies ($n_{\text{range}}=11\text{--}249$ trans women). The largest of these studies reported that grip strength decreased by a mean of -1.8 kg in trans women following 12 months of GAHT⁷³; however, this study did not include a comparison to a cisgender cohort. A smaller longitudinal study ($n=20$) of the same time frame found no change in grip strength following 12 months GAHT in trans women,⁶⁷ and a more recent 2021 prospective study ($n=11$) found a reduction of 7.7% in grip strength over a shorter 6 month time frame.⁶⁹ Conversely, a 2020 longitudinal study reported

that muscle strength, determined using isokinetic dynamometry, was maintained in trans women ($n=11$) following 12 months, with muscle strength greater in trans women, compared to height-adjusted cis women ($n=14$).⁵² Limitations of this study were in its small sample size, with differences in age and body size between groups not adjusted for. Furthermore, the study and the 2020 systematic review were limited to 12 months, with potential further reductions in strength after this time was not captured. Cross-sectional analyses of trans women ($n=23$) using GAHT for longer than 12 months (mean duration of 8 years), found that trans women had lower peak quadriceps strength (-25%), peak bicep strength (-33%) and grip strength (-23%) compared to height-matched cis men.⁷⁴ However, this study did not include comparisons with matched cis women and did not match participants for physical activity level or training status. In addition to the limitations mentioned above, another limitation of prior studies is that the majority of studies are conducted in untrained trans women. A further cross-sectional study examining muscle strength using a hand grip dynamometer in trans women who had completed an average of 14.1 ± 3.5 years GAHT found that, on average, the muscle strength data of trans women fell below that of cisgender men but was significantly greater (16%) than the values of cisgender women.⁷⁵ Most studies assessing muscle strength have utilized grip strength which has not been adjusted for hand size and overall is unlikely to be a good reflection of overall physical performance. Further, only two related studies have examined athletic performance in athletic trans people. A 2021 retrospective analysis of fitness tests performed by the United States military commencing GAHT during service was the first study to examine GAHT-related changes in trained trans individuals.⁷⁶ Before beginning GAHT, military trans women ($n=46$) performed 31% more push-ups and 15% more sit-ups in 1 min compared to cis women ($n>500,000$), with these differences disappearing after 2 years of GAHT.⁷⁶ A later study that drew from a larger sample set of US military personnel ($n=228$ trans women) with extended follow up, found similar trends for the sit-up and push-up performance of trans women to decline steadily alongside months using GAHT.⁷⁷ However, compared to cisgender women, trans women performed significantly better in sit-ups and push-ups, up until 4 years of GAHT, where no statistically significant

difference was found in sit-up performance between trans and cisgender women.⁷⁷ Differences in findings across studies likely stem from differences in sample size and dropout and should be interpreted with caution given the large number of dropouts across both studies. Collectively, research is limited with few longitudinal studies beyond 2 years and a lack of prospectively recruited comparison groups. Results suggest a reduction in muscle strength after starting GAHT, but the duration to see significant differences is unclear.

In trans men, limited studies have examined the effect of GAHT on muscle strength measures. A prospective study reported that knee extension and flexion torque, which is the maximum force applied by a muscle that causes rotation, increased by 16% and 26%, respectively, in trans men ($n=12$) following 12 months GAHT,⁵² with further studies finding significant increases (~18%) in the grip strength of trans men over 12 months.^{53,73} A more recent 2022 cross-sectional study compared the hand grip strength of trans men undergoing GAHT for 10–28 months ($n=19$), and cis men ($n=19$) and demonstrated that cis men had 40% greater hand grip strength than that of trans men.³⁵ Further, a retrospective analysis of fitness tests performed by the United States military found that trans men ($n=29$) performed fewer push-ups (43%) before GAHT commencement, compared to cis men ($n<2.3$ million), with no differences between groups found following 1 year of GAHT. This study also reported that trans and cis men were comparable in sit-up performance prior to GAHT commencement, with the sit-up performance of trans men then exceeding that of cisgender males after 12 months of GAHT.⁷⁶ A later study in an extended sample set of US military ($n=146$ trans men⁷⁷), demonstrated that trans men had lower performance in the push-up tests (–10%) compared to cisgender men, up to 2 years of GAHT, with greater performance in these tests, compared to cisgender, shown following 3 (+5%) and 4 (+15%) years GAHT.⁷⁷ This study also reported better performance in the sit-up test at 1 year GAHT (+4%), relative to cisgender men, which rose to as high as 13% following 4 years GAHT.⁷⁷ However, limitations of these studies include high variability in the time between baseline and follow-up periods in trans participants, fitness tests being performed in a non-controlled setting, and factors such as training volume, intensity or type

were not accounted for.⁷⁶ Further research addressing limitations is required to improve understanding of strength differences between trans and cis men and provide updated guidance for participation in high-impact sports that rely on muscle mass and strength.

Alterations in muscle at a cellular or tissue level may partially underpin observed physiological changes or differences in muscle strength between cisgender individuals and trans individuals using GAHT. In this regard, the theory of ‘muscle memory’⁷⁸ has been increasingly discussed in relation to trans women athletes. Muscle memory theorizes that muscle retains the capacity to perform tasks it has previously undergone, with suggestions that trans women, therefore, may retain muscle strength advantages over cis women after transition due to cellular or epigenetic marks retained from prior life exposures to testosterone and myonuclei retention. However, it is important to highlight that this area has not been explored in those undergoing GAHT. Myonuclei retention plays a potential role in muscle memory,⁷⁹ with prior research showing myonuclei numbers are associated with training and testosterone use.^{79,80} In mice, a 2010 study in mice⁷⁹ found muscle fibre myonuclei increased after exercise, with the acquired myonuclei retained during the detraining phase. Similar findings were found in female mice treated with testosterone propionate, with a 66% increase in myonuclei content observed following 14 days and a 10% elevation in myonuclei remaining across the mouse lifespan.⁸¹ In humans, a 1999 study compared the muscle physiology of powerlifters who were users ($n=9$) and non-users of high-dose steroids ($n=10$) and found that the number of myonuclei and the proportion of myonuclei were significantly higher in the steroid users.⁸⁰ However, the concept of muscle memory has mainly been based on animal models, with a recent comprehensive review⁸² and systematic review and meta-analysis⁸³ finding no clear evidence for the existence of muscle memory by myonuclei retention in both animal and human studies. In particular, it appears that ‘muscle memory’ is a phenomenon that is not observed in many studies or not retained long term, with many studies,^{82,83} and it could be debated that any memory marks altered during previous training would be easily modified or overridden by changes in sex hormone concentrations and the subsequent systematic effects of GAHT. Further

research exploring physiological and epigenetic changes in trans men and women using GAHT could provide great molecular insights into the performance differences between cis and trans populations and is an area of future need.

Effects of GAHT on endurance performance measures

Little research to date has explored the effects of GAHT on endurance performance. Thus far, the evidence indicates that GAHT can influence markers of oxygen transport, such as haemoglobin.⁵⁷ The oxygen-carrying capacity of haemoglobin plays a vital role in determining endurance performance, with maximal oxygen uptake ($\text{VO}_{2\text{max}}$), the most robust measure of endurance performance, increasing by approximately 1% for each 3 g/L increase in haemoglobin.⁸⁴ To date, one study has been performed assessing the effects of feminizing GAHT on $\text{VO}_{2\text{peak}}$, with a cross-sectional study finding that trans women ($n=15$) who had received GAHT for 14.1 ± 3.5 years, on average, had a $\text{VO}_{2\text{peak}}$ /weight (relative to bodyweight) that fell lower the mean $\text{VO}_{2\text{peak}}$ /weight for both cisgender men and cisgender women. However, significance was only found between trans women and cisgender men, not trans women and cisgender women.⁷⁵ This study also found that for absolute $\text{VO}_{2\text{peak}}$, trans women $\text{VO}_{2\text{peak}}$ was significantly higher than that of cisgender women (20%), and significantly lower than that of cisgender men (29%).⁷⁵ A limitation of this study is in its cross-sectional design, with the $\text{VO}_{2\text{peak}}$ of individuals before GAHT commencement not known, and the degree of change in $\text{VO}_{2\text{peak}}$ following GAHT commencement not able to be investigated. A 2021 systematic review that evaluated nine haemoglobin-related longitudinal studies ($n_{\text{range}} = 12\text{--}239$ trans women) and six cross-sectional studies ($n_{\text{range}} = 23\text{--}182$ trans women) reported consistent reductions in circulating haemoglobin concentrations (-3.4 to -14%) following 3–36 months of GAHT in trans women.⁵⁷ Findings of this systematic review further demonstrated that haemoglobin concentrations of trans women decreased to typical female ranges after 4 months GAHT.⁵⁷ In trans men using GAHT for at least 12 months ($n=79$), a cross-sectional study found that haemoglobin concentrations (12.8–17.4 mg/dL) were comparable to cis male reference ranges (13.0–18.0 mg/dL).⁸⁵

Findings of GAHT-related changes in haemoglobin suggest gender-affirming hormones influence oxygen transport and uptake in trans individuals, with likely broader impacts on endurance performance. However, no studies have yet reported changes in robust and direct markers of endurance fitness (e.g. $\text{VO}_{2\text{max}}$). A retrospective study of military personnel, previously described in the above muscle strength section, compared 1.5 mi run times between trans men ($n=49$), trans women ($n=46$) and cis military personnel not undergoing hormone therapy.⁷⁶ This study reported that trans women had 21% faster running times than cis women before hormone commencement, with this difference reducing to 9% after 2 years of GAHT.⁷⁶ No differences in running times were observed between trans men and cis men.⁷⁶ A more recent study, which analysed a larger data set of this US military sample ($n=228$ trans women, 146 trans men), with extended follow-up, found trends for 1.5 mi run times to decrease with GAHT in trans women, with run times statistically greater than cisgender women up until 2 years of GAHT. This study also found that, compared to cisgender men, trans men had statistically lower run times up until 3 years into GAHT.⁷⁷ Differences in findings across these US military studies likely arises from differences in sample size and dropout, with a limitation of these studies being the large decrease in participants over the respective timelines, with the more recent study by Chiccarelli *et al.*⁷⁷ seeing a 96% drop-off from baseline. However, as touched on in previous sections, these studies also have limitations in being retrospective analyses of fitness tests performed in a non-controlled environment, with confounders not considered in analyses. Overall, current data indicates haemoglobin concentrations are comparable between trans women and cis men, as well as trans men and cis men following several months of GAHT. It is unclear whether changes in haemoglobin reflect changes in endurance performance, with a clear lack of studies exploring this area. A focus of further research should be on evaluating potential changes in $\text{VO}_{2\text{max}}$ in trans individuals using GAHT, given that $\text{VO}_{2\text{max}}$ is a major predictor of performance in sports such as distance cycling and distance running,^{37,86,87} with clear sex differences in this measure reported.³⁶ There is a clear lack of control for confounders in this area, with factors such as body size, muscle mass and fat mass not considered during analysis in previous studies.

Effects of GAHT on bone health

Markers of bone health, such as BMD, may be influenced by GAHT, with implications for participation in weight-bearing sports. Bone has key functions in structural support and allowing efficient movement,⁸⁸ with mechanical loading involved in physical activity has a role in bone development.⁸⁹ Athletes who compete in weight-bearing sports, such as gymnastics, are found to have 5–15% higher BMD than non-athletes.⁹⁰ Furthermore, high-impact activities, such as endurance running, are associated with a greater risk of bone stress fractures.⁹¹ This highlights the importance of bone health and strength for performance and safety in a range of sports.

A systematic review and meta-analysis conducted in 2019 ($n=812$ trans women, $n=487$ trans men), found no effect of GAHT on BMD at all sites evaluated (femoral neck, total femur and lumbar spine), except for the lumbar spine BMD for trans women, where a small but significant increase was observed after 12 and ≥ 24 months of GAHT.⁹² However, this review concluded that current evidence is of low to moderate quality, with inconsistencies commonly reported across studies comparing BMD changes in trans and cis cohorts. For example, A 6-month prospective study in trans women ($n=11$) reported lumbar spine BMD significantly increased by 3.9%, but changes in femoral neck and total femur BMD were not significant, and comparisons to cisgender cohorts were not made.⁶⁹ A more recent 2022 cross-sectional study reported total BMD and femoral neck BMD was lower in trans men using GAHT for at least 6 months ($n=19$), when compared to age and BMI-matched cis men ($n=19$).⁶³ In contrast, another cross-sectional study in trans men ($n=15$) and women ($n=24$) using GAHT for 2–25 years found BMD Z scores for trans men did not differ from age-matched cis male reference groups, except for distal tibial diaphysis BMD, which was lower in trans men.⁹³ This study further reported whole body and distal tibial diaphysis BMD Z scores were higher in trans women compared to cis female reference values.⁹³ Unclear findings likely reflect small sample sizes and studies having differing hormone regimes, durations and comparator groups. Furthermore, these studies were limited in assessing areal BMD using dual-energy X-ray absorptiometry, which does not provide information on volumetric and structural changes. Addressing this limitation, a cross-sectional study published in 2022 examined

differences in bone microarchitecture and volumetric BMD (via high-resolution peripheral quantitative computed tomography images) between trans and cis cohorts matched by age (trans men: $n=41$, trans women: $n=51$, cis women: $n=71$, cis males: $n=51$). This study found trans men had significantly higher total volumetric BMD (+0.63 SD) and greater trabecular thickness (+0.38 SD) than cis women. Relative to cis males, trans women had lower total (−0.68 SD) and cortical volumetric BMD (−0.70 SD), and higher cortical porosity (+0.70 SD). Findings suggest trans women experience deteriorated bone microarchitecture with GAHT, with bone microarchitecture not compromised in trans men.⁹⁴ However, additional prospective longitudinal studies are needed to confirm findings. Given BMD correlates with measures of muscle strength in weight-bearing sports and fracture risk in endurance sports,^{90,91} this is an area requiring further investigation and consideration in sports guidelines.

Effects of GAHT on metabolic and disease risk markers

Gender-affirming hormones may influence risk factors for cardiometabolic diseases, such as blood lipids and insulin resistance.^{28,95} This is of relevance to sports participation at both elite and community levels, given regular participation in sports and exercise plays an important role in improving and mediating cardiometabolic health.⁹⁶

An analysis of United States survey data ($n \sim 2.7$ million adults, including ~ 3000 trans adults) reports a two to fourfold increase in the rate of myocardial infarctions in both trans men and women, compared to cis population,⁹⁷ with a further cross-sectional analysis of ~ 4000 trans adults and cis reference cohorts ($n \sim 100,000$) finding a significantly higher incidence of venous thromboembolism in trans women.⁹⁸ A 2019 systematic review²⁸ found varied evidence of GAHT-related changes in cardiometabolic risk markers, including high-density lipoprotein (HDL) concentrations and low-density lipoprotein (LDL) concentrations. In studies of trans women, increases in HDL concentrations were reported (14–27% over 6–24 months GAHT), with decreases in LDL concentrations (8–13% from 6 to 18 months GAHT).^{99–102} In trans men, HDL concentrations increased in several studies (3–33%

over 6–24 months of GAHT), but LDL findings were inconclusive.^{99–104}

A systematic review of 26 studies ($n=751$ trans men, $n=689$ trans women) reported that masculinizing GAHT has no impact on insulin resistance in trans men but that feminizing GAHT may exacerbate insulin resistance in trans women.⁹⁵ However, a 2022 retrospective study compared the incidence of type 2 diabetes in adult trans women ($n=2585$) and trans men ($n=1514$) with the general population rates from their birth-assigned sex and found no significant difference in the incidence of diabetes.¹⁰⁵ Therefore, while the current evidence suggests GAHT may impact insulin resistance, this does not seem to confer a greater risk of type 2 diabetes. Additional long-term controlled studies are required, with current evidence largely based on uncontrolled studies of short duration (<2 years) and findings likely confounded by lifestyle factors such as physical inactivity²³ and alcohol and nicotine use¹⁰⁶ in trans individuals.

Limitations and future directions

Current knowledge on the sports implications of gender-affirming hormone therapies is lacking (Figure 1). This is an emerging field of growing interest in health and sporting fields, with clear limitations in the existing literature. There is a clear lack of long-term longitudinal studies ($2+$ years), and therefore limited data and information on the long-term performance changes in both trans men and trans women using GAHT. Further, many studies to date have been cross-sectional or uncontrolled cohort studies that do not include accurate comparator groups of cis participants matched for important attributes such as age, height and BMI. Matching for these factors in research studies is needed to more robustly evaluate the effect of GAHT on the measures of interest, separate to possible confounding factors, and to subsequently avoid over and underestimating the influence of GAHT. GAHT-related changes in testosterone concentrations and body composition are the most researched areas, with these measures not perfect indicators of athletic capacity. Current evidence covers a limited range of performance measures (i.e., muscle strength), with future directions in investigating important markers of fitness (e.g. VO_2max , cardiac output, maximal power output, W_{max} /body weight ratio) and muscle function

(e.g. epigenetics, muscle enzyme activity, fibre type). Of relevance is a well-controlled prospective matched cohort study (the GAME Study) just commenced by the authors that aim to address this limitation and examine a wider range of performance and muscle measures.¹⁰⁷ Further, it is notable for highlighting that no studies exist comparing fitness and performance measures between trans individuals who commenced before and after the onset of puberty, and the related effects of puberty blockers; this is an area of future research, given recent policies made by FINA to exclude trans people from competing in elite competitions, unless they began GAHT pre-puberty,²⁰ it is of interest to understand the effects of puberty blockers on measures of sports performance. Importantly, most studies have been conducted in untrained trans cohorts, with it was unclear whether findings are generalizable to athletes. The 2021 IOC framework on Fairness, Inclusion and Non-Discrimination on the Basis of Gender Identity and Sex Variations⁸ states that sporting guidelines should be based on robust and peer-reviewed research and largely based on data gathered for groups with appropriate athletic engagement and measures appropriate to specific sport disciplines. This is not the current level of evidence available, and limitations in current knowledge should therefore be highlighted in current sporting guidelines and policy actions to prevent stigma, discrimination and potential harm to the trans community. Sporting organizations and federations should be cautious in implementing exclusive policies, particularly with a lack of trans people at the elite level, and in the absence of evidence, an individualized approach to the assessment of trans people should be taken.

Conclusion

The present literature indicates feminizing GAHT effectively lowers testosterone concentrations to the typical female ranges in trans women after 12 months, with masculinizing hormone therapies increasing concentrations to the male range in trans men. Feminizing GAHT is further shown to increase fat mass and decrease lean mass in trans women, with opposite effects observed in trans men using masculinizing GAHT. In trans women, muscle strength measured in small studies is shown to decrease or not change following at least 12 months of GAHT and appears to continue to decrease beyond 12 months, with increased strength observed in

trans men using GAHT. Longer durations of GAHT appear to be associated with significantly lower muscle strength compared to cis men. Studies comparing strength measures between trans and cis women report unclear results, which are dependent on study design, sample size and the strength outcomes measured. Little research has explored the effects of GAHT on endurance performance, but from the current literature, endurance performance is impaired in trans women (based on 1.5 mi run and VO_2max data), and is somewhat improved in trans men (based on 1.5 mi run time data). Further, consistent changes in markers of oxygen transport (haemoglobin) suggest GAHT influences oxygen capacity; however, further investigations are required. Prior research reports varied GAHT-related changes in HDL, LDL and insulin resistance, with data limited due to a lack of long-term longitudinal studies and likely confounded by lifestyle factors. Overall, current limitations of this field include a lack of long-term studies, small sample sizes and the use of no or inadequate comparison groups. Future directions should investigate a broader range of markers related to fitness, and muscle health and functioning. There is a need for well-controlled longitudinal studies on GAHT in non-athletic and athletic trans people to address the current limitations and provide more robust data to inform inclusive and fair sporting programmes, policies, and guidelines.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Author contribution(s)

Ethan Moreland: Conceptualization; Data curation; Investigation; Methodology; Writing – original draft; Writing – review & editing.

Ada S. Cheung: Conceptualization; Investigation; Supervision; Writing – original draft; Writing – review & editing.

Danielle Hiam: Investigation; Writing – original draft; Writing – review & editing.

Brendan J. Nolan: Conceptualization; Investigation; Writing – original draft; Writing – review & editing.

Shanie Landen: Conceptualization; Investigation; Writing – original draft; Writing – review & editing.

Macsue Jacques: Conceptualization; Investigation; Writing – original draft; Writing – review & editing.

Nir Eynon: Conceptualization; Investigation; Supervision; Writing – original draft; Writing – review & editing.

Patrice Jones: Conceptualization; Investigation; Methodology; Supervision; Writing – original draft; Writing – review & editing.

Acknowledgements

None.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: NE is supported by a National Health & Medical Research Council (NHMRC) Career Development Fellowship (#1140644), ASC is supported by an NHMRC Investigator Grant (#2008956) and BJN is supported by an NHMRC Postgraduate Research Scholarship (#2003939).

Competing interests

The authors declare that there is no conflict of interest.

Availability of data and materials

Not applicable.

ORCID iDs

Ada S. Cheung  <https://orcid.org/0000-0002-7158-8804>

Patrice Jones  <https://orcid.org/0000-0002-4948-1445>

References

- Fontanari AMV, Vilanova F, Schneider MA, *et al.* Gender Affirmation is associated with transgender and Gender Nonbinary Youth Mental Health Improvement. *LGBT Health* 2020; 7: 237–247.
- Hughto JMW, Gunn HA, Rood BA, *et al.* Social and medical gender affirmation experiences are inversely associated with mental health problems in a U.S. non-probability sample of transgender adults. *Arch Sex Behav* 2020; 49: 2635–2647.

3. Glynn TR, Gamarel KE, Kahler CW, *et al.* The role of gender affirmation in psychological well-being among transgender women. *Psychol Sex Orient Gender Divers* 2016; 3: 336–344.
4. Coleman E, Bockting W, Botzer M, *et al.* Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgend* 2012; 13: 165–232.
5. Wylie K, Barrett J, Besser M, *et al.* Good practice guidelines for the assessment and treatment of adults with gender Dysphoria. *Sex Relation Ther* 2014; 29: 154–214.
6. Giltay EJ and Gooren LJ. Effects of sex steroid deprivation/administration on hair growth and skin sebum production in transsexual males and females. *J Clin Endocrinol Metab* 2000; 85: 2913–2921.
7. Wierckx K, Mueller S, Weyers S, *et al.* Long-term evaluation of Cross-Sex hormone treatment in transsexual persons. *J Sex Med* 2012; 9: 2641–2651.
8. International Olympic Committee. IOC framework on fairness, inclusion and non-discrimination on the basis of gender identity and sex variations, https://stillmed.olympics.com/media/Documents/News/2021/11/IOC-Framework-Fairness-Inclusion-Non-discrimination-2021.pdf?_ga=2.145205876.1586162731.1653964229-1319763856.1653964229 (2021, accessed 1 May 2022).
9. Sandbakk Solli GS and Holmberg HC. Sex differences in world-record performance: the influence of sport discipline and competition duration. *Int J Sports Physiol Perform* 2018; 13: 2–8.
10. Santisteban KJ, Lovering AT, Halliwill JR, *et al.* Sex differences in VO(2max) and the impact on endurance-Exercise Performance. *Int J Environ Res Public Health* 2022; 19: 4946.
11. Hallam LC and Amorim FT. Expanding the gap: an updated look into Sex differences in running performance. *Front Physiol* 2021; 12: 804149.
12. Safer JD. Fairness for transgender people in Sport. *J Endocr Soc* 2022; 6: bvac035.
13. Goodman M, Adams N, Corneil T, *et al.* Size and distribution of transgender and gender nonconforming populations: a narrative review. *Endocrinol Metab Clin North Am* 2019; 48: 303–321.
14. Anderson E and Travers A. *Transgender athletes in competitive sport*. London: Taylor & Francis, 2017.
15. Buzuvis E. Sexual Minorities in sports: Prejudice at Play. In: Sartore-Baldwin Melanie L. (ed.) *Transsexual and intersex athletes*. Boulder: Lynne Rienner Publishers, 2013, pp. 55–71.
16. Scovel S, Nelson M and Thorpe H. Media framings of the transgender athlete as “Legitimate Controversy”: the Case of Laurel Hubbard at the Tokyo Olympics. *Commun Sport* 2022: 21674795221116884.
17. Ennis D. Chris Mosier makes history at Olympic trials, calling it ‘incredible and heartbreaking’ SB Nation Outsports, <https://www.outsports.com/2020/1/27/21083208/chris-mosier-transgender-olympic-trials-tokyo-2020-race-walking> (2020, accessed 1 May 2022).
18. Steele L. Chris Mosier on Making History as First Trans Member of Team USA: RollingStone 2016, <https://www.rollingstone.com/culture/culture-sports/chris-mosier-on-making-history-as-first-trans-member-of-team-usa-250971/> (2016, accessed 1 February 2022).
19. International Olympic Committee. IOC Consensus Meeting on Sex Reassignment and Hyperandrogenism 2015, https://stillmed.olympic.org/Documents/Commissions_PDFfiles/Medical_commission/2015-11_ioc_consensus_meeting_on_sex_reassignment_and_hyperandrogenism-en.pdf (2015, accessed 5 September 2021).
20. International Swimming Federation (FINA). Policy on eligibility for the men’s and women’s competition categories 2022, <https://resources.fina.org/fina/document/2023/03/27/dbc3381c-91e9-4ea4-a743-84c8b06debef/Policy-on-Eligibility-for-the-Men-s-and-Women-s-Competition-Categories-Version-on-2023.03.24.pdf> (2022, accessed 1 February 2023).
21. Bikes SN. Australian sports set guidelines for inclusion of trans athletes Reuters Sports News2020, <https://www.reuters.com/article/us-sports-australia-transgender-idUSKBN26M42M> (2020, accessed 1 May 2022).
22. Human Rights Campaign Foundation. Play to Win: Improving the Lives of LGBTQ Youth in Sports 2017, <https://assets2.hrc.org/files/assets/resources/PlayToWin-FINAL.pdf> (2017, accessed 15 March 2022).
23. Jones BA, Haycraft E, Bouman WP, *et al.* The levels and predictors of physical activity engagement within the treatment-seeking transgender population: a matched control study. *J Phys Act Health* 2018; 15: 99–107.

24. López-Cañada E, Devis-Devis J, Valencia-Peris A, *et al.* Physical Activity and sport in trans Persons before and after gender disclosure: prevalence, frequency, and type of activities. *J Phys Act Health* 2020; 17: 650–656.
25. Braumüller B, Menzel T and Hartmann-Tews I. Gender Identities in organized sports-athletes' experiences and organizational strategies of inclusion. *Front Sociol* 2020; 5: 578213.
26. Veliz P, Snyder M and Sabo D. *The state of high school sports in America: an Evaluation of the Nation's most popular extracurricular activity.* New York, NY: Women's Sports Foundation, 2019.
27. Dhejne C, Van Vlerken R, Heylens G, *et al.* Mental health and gender dysphoria: a review of the literature. *Int Rev Psychiatry* 2016; 28: 44–57.
28. Defreyne J, Van de Bruaene LDL, Rietzschel E, *et al.* Effects of gender-affirming hormones on lipid, metabolic, and cardiac surrogate blood markers in transgender persons. *Clin Chem* 2019; 65: 119–134.
29. Cheuvront SN, Carter R, Deruisseau KC, *et al.* Running performance differences between men and women: an update. *Sports Med* 2005; 35: 1017–1024.
30. Seiler S, De Koning JJ and Foster C. The fall and rise of the gender difference in elite anaerobic performance 1952–2006. *Med Sci Sports Exerc* 2007; 39: 534–540.
31. Handelsman DJ. Sex differences in athletic performance emerge coinciding with the onset of male puberty. *Clin Endocrinol* 2017; 87: 68–72.
32. Nikolaidis PT, Rosemann T and Knechtle B. Sex differences in the age of Peak marathon race time. *Chin J Physiol* 2018; 61: 85–91.
33. Helgerud J, Ingjer F and Strømme SB. Sex differences in performance-matched marathon runners. *Eur J Appl Physiol Occup Physiol* 1990; 61: 433–439.
34. Sparling PB, O'Donnell EM and Snow TK. The gender difference in distance running performance has plateaued: an analysis of world rankings from 1980 to 1996. *Med Sci Sports Exerc* 1998; 30: 1725–1729.
35. Chen G, Liu L and Yu J. A comparative study on strength between American college male and female students in Caucasian and Asian populations. *Sport Sci Rev* 2012; 21: 153–165.
36. Kenney WL, Wilmore JH and Costill DL. *Physiology of sport and exercise.* United States of America: Human Kinetics, 2021.
37. Foster C. VO2 max and training indices as determinants of competitive running performance. *J Sports Sci* 1983; 1: 13–22.
38. Stone MH, Sands WA, Pierce KC, *et al.* Relationship of maximum strength to weightlifting performance. *Med Sci Sports Exerc* 2005; 37: 1037–1043.
39. Schorr M, Dichtel LE, Gerweck AV, *et al.* Sex differences in body composition and association with cardiometabolic risk. *Biol Sex Differ* 2018; 9: 28.
40. Lunt M, Felsenberg D, Reeve J, *et al.* Bone density variation and its effects on risk of vertebral deformity in men and women studied in thirteen European centers: the EVOS Study. *J Bone Miner Res* 1997; 12: 1883–1894.
41. Bacon WE. Secular trends in hip fracture occurrence and survival: age and sex differences. *J Aging Health* 1996; 8: 538–553.
42. Nieves JW, Formica C, Ruffing J, *et al.* Males have larger skeletal size and bone mass than females, despite comparable body size. *J Bone Miner Res* 2005; 20: 529–535.
43. Auyeung TW, Lee JS, Kwok T, *et al.* Testosterone but not estradiol level is positively related to muscle strength and physical performance independent of muscle mass: a cross-sectional study in 1489 older men. *Eur J Endoc* 2011; 164: 811–817.
44. Clark RV, Wald JA, Swerdloff RS, *et al.* Large divergence in testosterone concentrations between men and women: Frame of reference for elite athletes in sex-specific competition in sports, a narrative review. *Clin Endocrinol* 2019; 90: 15–22.
45. Herbst KL and Bhasin S. Testosterone action on skeletal muscle. *Curr Opin Clin Nutr Metab Care* 2004; 7: 271–277.
46. Alexander SE, Pollock AC and Lamon S. The effect of sex hormones on skeletal muscle adaptation in females. *Eur J Sport Sci* 2022; 22: 1035–1045.
47. Moore E, Wisniewski A and Dobs A. Endocrine treatment of transsexual people: A review of treatment regimens, outcomes, and adverse effects. *J Clin Endocrinol Metab* 2003; 88: 3467–3473.
48. Angus LM, Nolan BJ, Zajac JD, *et al.* A systematic review of antiandrogens and feminization in transgender women. *Clin Endocrinol* 2021; 94: 743–752.

49. Angus L, Leemaqz S, Ooi O, *et al.* Cyproterone acetate or spironolactone in lowering testosterone concentrations for transgender individuals receiving oestradiol therapy. *Endocr Connect* 2019; 8: 935–940.
50. Gava G, Cerpolini S, Martelli V, *et al.* Cyproterone acetate vs leuprolide acetate in combination with transdermal oestradiol in transwomen: a comparison of safety and effectiveness. *Clin Endocrinol* 2016; 85: 239–246.
51. Kuipers SM, Wiepjes CM, Conemans EB, *et al.* Toward a lowest effective dose of cyproterone acetate in trans women: Results from the ENIGI study. *J Clin Endocrinol Metab* 2021; 106: e3936–e45.
52. Wiik A, Lundberg TR, Rullman E, *et al.* Muscle strength, size, and composition following 12 months of gender-affirming treatment in transgender individuals. *J Clin Endocrinol Metab* 2020; 105: e805–e13.
53. Van Caenegem E, Wierckx K, Taes Y, *et al.* Body composition, bone turnover, and bone mass in trans men during testosterone treatment: 1-year follow-up data from a prospective case-controlled study (ENIGI). *Eur J Endoc* 2015; 172: 163–171.
54. Gava G, Mancini I, Cerpolini S, *et al.* Testosterone undecanoate and testosterone enanthate injections are both effective and safe in transmen over 5 years of administration. *Clin Endocrinol* 2018; 89: 878–886.
55. Pelusi C, Costantino A, Martelli V, *et al.* Effects of three different testosterone formulations in female-to-male transsexual persons. *J Sex Med* 2014; 11: 3002–3011.
56. World Rugby. *World Rugby transgender guidelines 2020*, <https://www.world.rugby/the-game/player-welfare/guidelines/transgender#SummaryforTransgenderWomen> (2020, accessed 15 March 2022).
57. Harper J, O'Donnell E, Sorouri Khorashad B, *et al.* How does hormone transition in transgender women change body composition, muscle strength and haemoglobin? Systematic review with a focus on the implications for sport participation. *Br J Sports Med* 2021; 55: 865–872.
58. Klaver M, de Blok CJM, Wiepjes CM, *et al.* Changes in regional body fat, lean body mass and body shape in trans persons using cross-sex hormonal therapy: results from a multicenter prospective study. *Eur J Endoc* 2018; 178: 163–171.
59. Suppakitjanusant P, Ji Y, Stevenson MO, *et al.* Effects of gender affirming hormone therapy on body mass index in transgender individuals: a longitudinal cohort study. *J Clin Transl Endocrinol* 2020; 21: 100230.
60. Shadid S, Abosi-Appeadu K, De Maertelaere A-S, *et al.* Effects of gender-affirming hormone therapy on insulin sensitivity and incretin responses in transgender people. *Diabetes Care* 2020; 43: 411–417.
61. Fisher AD, Castellini G, Ristori J, *et al.* Cross-Sex hormone treatment and psychobiological changes in transsexual persons: Two-Year Follow-Up data. *J Clin Endocrinol Metab* 2016; 101: 4260–4269.
62. Berra M, Armillotta F, D'Emidio L, *et al.* Testosterone decreases adiponectin levels in female to male transsexuals. *Asian J Androl* 2006; 8: 725–729.
63. Andrade SRL, Mucida YM, Xavier JDC, *et al.* Bone mineral density, trabecular bone score and muscle strength in transgender men receiving testosterone therapy versus cisgender men. *Steroids* 2022; 178: 108951.
64. van Velzen DM, Nota NM, Simsek S, *et al.* Variation in sensitivity and rate of change in body composition: steps toward individualizing transgender care. *Eur J Endocrinol* 2020; 183: 529–536.
65. Tack LJW, Craen M, Lapauw B, *et al.* Proandrogenic and antiandrogenic progestins in transgender youth: differential effects on body composition and Bone Metabolism. *J Clin Endocrinol Metab* 2018; 103: 2147–2156.
66. Wierckx K, Van Caenegem E, Schreiner T, *et al.* Cross-Sex Hormone Therapy in trans persons is safe and effective at Short-Time Follow-Up: results from the European network for the investigation of gender incongruence. *J Sex Med* 2014; 11: 1999–2011.
67. Auer MK, Cecil A, Roepke Y, *et al.* 12-months metabolic changes among gender dysphoric individuals under cross-sex hormone treatment: a targeted metabolomics study. *Sci Rep* 2016; 6: 37005.
68. Klaver M, de Mutsert R, Wiepjes CM, *et al.* Early hormonal treatment affects body composition and body shape in young transgender adolescents. *J Sex Med* 2018; 15: 251–260.
69. Yun Y, Kim D and Lee ES. Effect of Cross-Sex hormones on body composition, bone mineral density, and muscle strength in trans women. *J Bone Metab* 2021; 28: 59–66.
70. Gooren LJ and Bunck MC. Transsexuals and competitive sports. *Eur J Endocrinol* 2004; 151: 425–429.

71. Imboden MT, Swartz AM, Finch HW, *et al.* Reference standards for lean mass measures using GE dual energy x-ray absorptiometry in Caucasian adults. *PLoS One* 2017; 12: e0176161.
72. Kirk B, Bani Hassan E, Brennan-Olsen S, *et al.* Body composition reference ranges in community-dwelling adults using dual-energy X-ray absorptiometry: the Australian Body Composition (ABC) study. *J Cachexia Sarcopenia Muscle* 2021; 12: 880–890.
73. Scharff M, Wiepjes CM, Klaver M, *et al.* Change in grip strength in trans people and its association with lean body mass and bone density. *Endocr Connect* 2019; 8: 1020–1028.
74. Lapauw B, Taes Y, Simoens S, *et al.* Body composition, volumetric and areal bone parameters in male-to-female transsexual persons. *Bone* 2008; 43: 1016–1021.
75. Alvares LAM, Santos MR, Souza FR, *et al.* Cardiopulmonary capacity and muscle strength in transgender women on long-term gender-affirming hormone therapy: a cross-sectional study. *Br J Sports Med* 2022; 56: 1292–1298.
76. Roberts TA, Smalley J and Ahrendt D. Effect of gender affirming hormones on athletic performance in transwomen and transmen: implications for sporting organisations and legislators. *Br J Sports Med* 2020; 55: 577–583.
77. Chiccarelli E, Aden J, Ahrendt D, *et al.* Fit transitioning: When can transgender Airmen Fitness Test in their affirmed gender? *Mil Med* 2022; Oct 22: usac320.
78. Sharples AP, Stewart CE and Seaborne RA. Does skeletal muscle have an ‘epi’-memory? The role of epigenetics in nutritional programming, metabolic disease, aging and exercise. *Aging Cell* 2016; 15: 603–616.
79. Bruusgaard JC, Johansen IB, Egner IM, *et al.* Myonuclei acquired by overload exercise precede hypertrophy and are not lost on detraining. *Proc Natl Acad Sci USA* 2010; 107: 15111–15116.
80. Kadi F, Eriksson A, Holmner S, *et al.* Effects of anabolic steroids on the muscle cells of strength-trained athletes. *Med Sci Sports Exerc* 1999; 31: 1528–1534.
81. Egner IM, Bruusgaard JC, Eftestøl E, *et al.* A cellular memory mechanism aids overload hypertrophy in muscle long after an episodic exposure to anabolic steroids. *J Physiol* 2013; 591: 6221–6230.
82. Snijders T, Aussieker T, Holwerda A, *et al.* The concept of skeletal muscle memory: evidence from animal and human studies. *Acta Physiol* 2020; 229: e13465.
83. Rahmati M, McCarthy JJ and Malakoutinia F. Myonuclear permanence in skeletal muscle memory: a systematic review and meta-analysis of human and animal studies. *J Cachexia Sarcopenia Muscle* 2022; 13: 2276–2297.
84. Otto JM, Montgomery HE and Richards T. Haemoglobin concentration and mass as determinants of exercise performance and of surgical outcome. *Extrem Physiol Med* 2013; 2: 33.
85. Greene DN, McPherson GW, Rongitsch J, *et al.* Hematology reference intervals for transgender adults on stable hormone therapy. *Clin Chim Acta* 2019; 492: 84–90.
86. Williams CJ, Gurd BJ, Bonafiglia JT, *et al.* A multi-center comparison of O(2peak) trainability between interval training and moderate intensity continuous training. *Front Physiol* 2019; 10: 19.
87. Butts NK, Henry BA and Mclean D. Correlations between VO2max and performance times of recreational triathletes. *J Sports Med Phys Fitness* 1991; 31: 339–344.
88. Schtscherbyna A, Ribeiro BG and Maria Lucia F. Chapter 6 – Bone Health, bone mineral density, and sports performance. In: Bagchi D, Nair S and Sen CK (eds) *Nutrition and enhanced sports performance*. United States of America: Academic Press, 2019, pp.73–81, 2nd ed.
89. Skerry TM. Mechanical loading and bone: what sort of exercise is beneficial to the skeleton? *Bone* 1997; 20: 179–181.
90. Fehling PC, Alekel L, Clasey J, *et al.* A comparison of bone mineral densities among female athletes in impact loading and active loading sports. *Bone* 1995; 17: 205–210.
91. Bennell K, Matheson G, Meeuwisse W, *et al.* Risk factors for stress fractures. *Sports Med* 1999; 28: 91–122.
92. Figuera TM, Ziegelmann PK, Rasia da Silva T, *et al.* Bone mass effects of Cross-Sex hormone therapy in transgender people: updated systematic review and meta-analysis. *J Endocr Soc* 2019; 3: 943–964.
93. Ruetsche AG, Kneubuehl R, Birkhaeuser MH, *et al.* Cortical and trabecular bone mineral density in transsexuals after long-term cross-sex hormonal treatment: a cross-sectional study. *Osteoporos Int* 2005; 16: 791–798.
94. Bretherton I, Ghasem-Zadeh A, Leemaqz SY, *et al.* Bone microarchitecture in transgender adults: A cross-sectional study. *J Bone Miner Res* 2022; 37: 643–648.

95. Spanos C, Bretherton I, Zajac JD, *et al.* Effects of gender-affirming hormone therapy on insulin resistance and body composition in transgender individuals: a systematic review. *World J Diabetes* 2020; 11: 66–77.
96. Belanger MJ, Rao P and Robbins JM. Exercise, physical activity, and Cardiometabolic Health: Pathophysiologic Insights. *Cardiol Rev* 2022; 30: 134–144.
97. Alzahrani T, Nguyen T, Ryan A, *et al.* Cardiovascular Disease Risk factors and myocardial infarction in the transgender population. *Circ Cardiovasc Qual Outcomes* 2019; 12: e005597.
98. Getahun D, Nash R, Flanders WD, *et al.* Cross-sex hormones and acute cardiovascular events in transgender persons: a cohort study. *Ann Intern Med* 2018; 169: 205–213.
99. Deutsch MB, Bhakri V and Kubicek K. Effects of cross-sex hormone treatment on transgender women and men. *Obstet Gynecol* 2015; 125: 605–610.
100. Olson-Kennedy J, Okonta V, Clark LF, *et al.* Physiologic response to gender-affirming hormones among transgender youth. *J Adolesc Health* 2018; 62: 397–401.
101. Fernandez JD and Tannock LR. Metabolic effects of hormone therapy in transgender patients. *Endocr Pract* 2016; 22: 383–388.
102. Vita R, Settineri S, Liotta M, *et al.* Changes in hormonal and metabolic parameters in transgender subjects on cross-sex hormone therapy: a cohort study. *Maturitas* 2018; 107: 92–96.
103. McCredie RJ, McCrohon JA, Turner L, *et al.* Vascular reactivity is impaired in genetic females taking high-dose androgens. *J Am Coll Cardiol* 1998; 32: 1331–1335.
104. Jacobeit JW, Gooren LJ and Schulte HM. Long-acting intramuscular testosterone undecanoate for treatment of female-to-male transgender individuals. *J Sex Med* 2007; 4: 1479–1484.
105. van Velzen D, Wiepjes C, Nota N, *et al.* Incident diabetes risk is not increased in transgender individuals using hormone therapy. *J Clin Endocrinol Metab* 2022; 107: e2000–e7.
106. Grant JM, Mottet L, Tanis JE, *et al.* *National transgender discrimination survey report on health and health care*. Findings of a study by the National Center for Transgender Equality and the National Gay and Lesbian Task Force. National Center for Transgender Equality, 2010. https://cancer-network.org/wp-content/uploads/2017/02/National_Transgender_Discrimination_Survey_Report_on_health_and_health_care.pdf (2010, accessed 1 February 2021).
107. Jones PR, Voisin S, Nolan BJ, *et al.* Uncovering the effects of gender affirming hormone therapy on skeletal muscle and epigenetics: protocol for a prospective matched cohort study in transgender individuals (the GAME study). *BMJ Open* 2022; 12: e060869.

Visit SAGE journals online
[journals.sagepub.com/
home/tae](https://journals.sagepub.com/home/tae)

S SAGE journals



Scandinavian Journal of Public Health, 2011; 39: 35–43

ORIGINAL ARTICLE

Physical fitness in northern European youth: Reference values from the Latvian Physical Health in Youth Study

MELITA SAUKA¹, ILGA S. PRIEDITE¹, LUDMILA ARTJUHOVA¹, VIESTURS LARINS², GUNTARS SELGA³, ÖRJAN DAHLSTRÖM⁴ & TOOMAS TIMPKA⁴

¹Sports Medicine State Agency, Riga, Latvia, ²Latvian Academy of Sport Education, Riga, Latvia, ³Riga Stradins University, Riga, Latvia, and ⁴Section of Social Medicine and Public Health, Linköping University, Sweden

Abstract

Aims: Physical fitness has the potential to be used as a key health determinant in youth. The aim of this study was to establish age- and sex-stratified reference values for health-related physical fitness in Latvian school-age children and to identify notable sex differences. **Methods:** Physical fitness was assessed using the EUROFIT test battery. Data were gathered from schoolchildren aged between 6 and 17 years ($n=10,464$) by a medical team using standardised methods (EUROFIT battery). Fitness levels, stratified by chronological age and sex, were computed as mean $\pm$ SD. Each test was also analysed for differences between males and females in each age group. **Results:** Boys performed better than girls in muscular endurance and strength, cardiorespiratory endurance, and speed–agility fitness tests. Girls had better flexibility than boys ($p < 0.001$). There was also a more pronounced improvement in physical fitness scores with age in boys compared with girls. **Conclusions:** Physical fitness reference values were developed for Latvian children and adolescents. These reference values will permit comparisons between students during physical education lessons and provide a baseline against which progress in physical fitness among northern European youths can be compared.

Key Words: *Cardiorespiratory endurance, EUROFIT, flexibility, muscular endurance*

Background

Physical fitness is a set of attributes that people have or achieve that relate to their ability to perform physical activity [1]. Health-related physical fitness consists of those components of physical fitness that are associated with health. Health-related fitness is demonstrated by being able to perform daily activities with vigour and by traits and capacities associated with a low risk for the development of chronic diseases and premature death [2]. Health-related fitness is commonly measured by physiological factors such as cardiorespiratory endurance, muscular strength, balance, and flexibility [3].

Health-related fitness has been established as a key health determinant that can be measured more reliably than physical activity in adulthood [4], childhood, and adolescence [5–9]. There is evidence

that a low-level physical fitness in youth predisposes individuals to health problems in adult life. In a recent systematic review, it was concluded that there is strong evidence that high cardiorespiratory endurance is associated with a healthier cardiovascular profile later in life and that muscular strength improvements from childhood to adolescence are negatively associated with changes in overall adiposity [10]. In addition to a healthier cardiovascular profile, it has been shown that the physical fitness aspects related to speed and dynamic strength in adolescence are associated with bone mineral density and the accumulation of bone mass later in life [11,12].

The possible uses of standardised measurement of physical fitness in childhood include assessment of opportunities for promotion of physical activity at the population level and evidence-based individual

Correspondence: Melita Sauka, PhD, Sports Medicine State Agency, Ministry of Health, Raina bulvaris 27, Riga, LV-1050, Latvia. Fax: +371 67227306. E-mail: melita@latnet.lv

(Accepted 11 July 2010)

© 2011 the Nordic Societies of Public Health
DOI: 10.1177/1403494810380298



counselling and prescription of exercise [4,13,14]. However, these uses of health-related fitness testing require that the assessments are based on up-to-date reference values. The positioning of individuals and groups in percentiles allows not only straightforward interpretation of the progression of physical fitness over time but also identification of cross-sectional deviations in need of immediate investigation and intervention. The method is thereby applicable in educational and health-care settings.

Aims

The aim of this study was to establish age- and sex-stratified reference values for health-related physical fitness in Latvian school-aged children and to identify notable sex differences.

Methods

The European Physical Fitness Test battery EUROFIT [15] was used to assess physical fitness. The nine tests in the battery cover cardiorespiratory endurance, muscular endurance, muscular strength, flexibility, balance, and speed and agility. The measurement protocols for the EUROFIT battery were approved by the Health Statistics and Medical Technologies State Agency of Latvia. The study was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Participants

Over a 5-year study period, pupils aged between 6 and 17 years were recruited to the study from public schools throughout Latvia. Information about the study was presented annually at regional meetings for principals. All study participation was voluntary. A total of approximately 2000 pupils from a representative mix of urban and rural schools were accepted into the study each year. To avoid test learning bias among the participants, each school participated only once. Written informed consent was sought from parents before the tests. Only those children who had obtained parental approval were approached about participation. Final informed consent was obtained from the children before starting data collection.

Data collection

The data were gathered by a specially trained medical team from the Sports Medicine State Agency from April 2004 to July 2009. The data collection

consisted of physical fitness testing and anthropometric measurements.

Cardiorespiratory endurance was evaluated by means of a 20-m endurance shuttle run (ESR) test, with the test subject running back and forth over a distance of 20 m. The initial speed was 8.5 km/h, which was increased by 0.5 km/h every minute (1 min is equal to one stage). The time at which a child dropped out was the indicator of his or her cardiorespiratory endurance.

Muscular endurance is the ability of a muscle group to execute repeated contractions over time or to maintain a maximal voluntary contraction for a prolonged period of time. The endurance strength (functional strength) of the upper body was measured using a bent arm hang (BAH) test, which determines the maximum length of time a subject can remain suspended by the arms from a bar. Trunk strength (i.e. abdominal muscular endurance) was assessed using a sit-ups (SUP) test completed within half a minute. The total number of completed sit-ups performed within 30 s was counted.

We assessed static and explosive strength. Handgrip strength refers to the maximal isometric force that can be generated mainly by the hand and a forearm muscle (upper limb) involved in performing the handgrip, and was assessed with a handgrip (HGR) test. Lower limb explosive strength was assessed using a standing broad jump (SBJ) test. The explosive strength developed by the legs from a standing position was recorded as the maximum horizontal distance covered by jumping with the feet together.

Mobility of the trunk and hips assessed using the sit-and-reach (SAR) test, which is a reflection of overall flexibility. Flexibility was measured in a sitting position with feet placed flat against a box and the fingertips placed on the edge of the top plate. The subject attempted to reach forward as far as possible, keeping his or her knees straight, without jerking and with hands stretched out. The maximum distance reached with the fingers when bending forward while sitting on the floor was recorded.

Speed covered assessment of speed of movement/agility and speed of limb movement. Speed of movement/agility was assessed using a 10 × 5 m shuttle run (SHR) test. This test assessed the subjects' speed of movement, agility and coordination in an integrated fashion. The subjects ran back and forth five times along a 10-m track at the highest speed possible. At the end of each track section, the subjects deposited or picked up a sponge from a line on the floor. Therefore, the test permitted measurement not only of speed of displacement, but also of agility and coordination.

The plate tapping test (PLT) assessed the speed and the coordination of limb movement. Two discs are placed with their centres 60 cm apart on the table. The subject moved the preferred hand back and forth between the discs over the hand in the middle as quickly as possible. This action was repeated for 25 full cycles (50 taps).

The flamingo balance test assessed the ability to balance successfully on one leg. The subject stood on a beam (metal beam 50 cm long, 5 cm high, and 3 cm wide) with shoes removed. The number of falls in 60 s of balancing was recorded.

Statistical analysis

Physical fitness levels, stratified by chronological age groups, were computed as the mean $\pm$ SD. Eleven chronological age groups were used for the analyses, ranging from 6 years (6.0–6.9 years) to 17 years (17.0–17.9 years). The physical fitness variables were also subjected to an analysis of differences (one-way ANOVA) between males and females within each age group. All calculations were performed using SPSS v.14.0 software for Windows. The significance level was set at $p < 0.05$.

Results

Data from 10,464 children and adolescents (5507 males and 4957 females) were analysed. The overall rate of participation in the tests was 95%. In the oldest age groups (16–17 years), where the motivation for participating in physical education classes was lower, the drop-out rates were higher than in the younger age groups. In schools where medical staff worked in a space smaller than 20 m in length, the 20-m ESR test was not done. These circumstances led to a high proportion of non-participation (60–70%) in this test.

The mean physical fitness scores for Latvian school-age children were grouped according to chronological age and sex from 6 to 17 years of age (Table I). Sex differences (favouring boys) in muscular endurance, as measured by the BAH and SUP tests, were statistically significant from 8 to 17 years of age. Analysis of strength (HGR and SBJ tests) revealed similar statistically significant differences in all age groups. However, girls had better flexibility (SAR test) than boys, with statistically significant differences in all age groups. Boys had better cardiorespiratory endurance (ESR test) compared with girls from the age of 9 years, and better speed-agility (SHR test) from the age of 7 years.

Physical fitness: Composition of reference values 37

The computed reference values stratified by age and sex and expressed in percentiles from P₅ to P₉₅ for trunk strength (SUP test), cardiorespiratory endurance (ESR test), and speed/agility (SHR test) are displayed in Tables II–IV. Aptitude for muscular endurance, strength, and cardiorespiratory endurance was higher for boys than for girls (Figure 1). The P₉₅ percentile line in girls was equivalent to the P₅₀ percentile line in boys. There was also a more pronounced improvement in physical fitness scores for boys than for girls with increasing age.

Discussion

We have developed population-based reference values for physical fitness in northern European (Baltic States and Scandinavia) schoolchildren and adolescents of varying age. To be able to compare the physical fitness level between countries, we used a randomised subject selection procedure, standardised measurements, and test methods that have been assessed for reliability [16]. In a recent review, it was concluded that the ESR test in particular is a valid measure of cardiorespiratory endurance, the SBJ test is a valid test for lower body muscular strength, and the HGR test is a valid measure of upper body muscular strength [17]. Nevertheless, it still must be remembered that the percentile lines identified in this study do not represent cut-off thresholds for pathological test results predictive for disease in adulthood, and that the associations between physical fitness and health are complex. For instance, improvements in muscular endurance, speed, and agility rather than cardiorespiratory endurance have a positive effect on skeletal health [18].

In our Latvian data, we found that boys performed better than girls in tests for muscular endurance and strength, cardiorespiratory endurance, and speed of movement/agility, whereas girls performed better in the flexibility tests. These observations correspond to the results of the HELENA study performed in 2006–08, using pooled data from children aged 13–17 years in 10 European countries [19]. Comparing our data with the results from the HELENA study, we found compelling similarities between functional strength (Figure 1a) and static strength (Figure 1b). The differences in mean BAH test scores between the current study and the HELENA study were only 2.64 s for boys and 2.62 s for girls, and only –1.4 kg for boys and –0.30 kg for girls for the HGR test. However, other reference values, such as for explosive strength (Figure 1c) and flexibility (Figure 1d), showed

Table I. Physical fitness scores (mean $\pm$ SD) by age and gender of Latvian school-age children from EUROFIT battery.

Age (years)	Muscular endurance			Muscular strength		Flexibility	Cardio-respiratory endurance		Speed		Total body balance	
	Bent arm hang (BAH) test (s)	Sit-ups (SUP) test (repeats)	Handgrip (HGR) test (kg)	Broad jump (SBJ) test (cm)	Sit and reach (SAR) test (cm)	Endurance shuttle run (ESR) test (min) ^a	Shuttle run test (10 × 5) (SHR) (s)	Plate tapping (PLT) test (s)	Flamingo test (no instabilities)			
Boys (<i>n</i> = 5507)												
6	5.6 ± 6.1	12.2 ± 5.4	8.6 ± 2.6 ^b	107.3 ± 25.0 ^b	16.5 ± 5.0 ^b	–	26.0 ± 2.7 ^c	23.4 ± 3.8	12.05 ± 4.5			
7	7.2 ± 6.9 ^c	15.1 ± 4.7	10.8 ± 3.2 ^b	120.3 ± 21.3 ^b	15.6 ± 5.6 ^b	1.9 ± 1.0	24.6 ± 2.8 ^b	21.0 ± 10.1	13.8 ± 2.2			
8	8.7 ± 8.0 ^b	17.1 ± 4.8 ^b	12.7 ± 3.3 ^b	132.2 ± 22.6 ^b	15.8 ± 5.8 ^b	2.2 ± 1.2 ^c	23.7 ± 2.8 ^b	18.4 ± 6.3	13.5 ± 2.5 ^c			
9	10.1 ± 9.5 ^b	18.7 ± 4.9 ^b	14.4 ± 3.7 ^b	138.8 ± 22.8 ^b	14.6 ± 6.2 ^b	2.6 ± 1.5 ^b	22.9 ± 2.6 ^b	16.7 ± 7.8	12.6 ± 3.0			
10	11.6 ± 9.9 ^b	20.2 ± 4.8 ^b	16.4 ± 4.1 ^b	147.3 ± 24.1 ^b	14.5 ± 6.2 ^b	3.2 ± 1.7 ^b	22.2 ± 2.8 ^b	15.4 ± 7.4	12.2 ± 3.3			
11	12.4 ± 10.6 ^b	21.3 ± 4.5 ^b	18.5 ± 5.1 ^b	152.0 ± 24.9 ^b	13.7 ± 6.6 ^b	3.7 ± 1.9 ^b	21.8 ± 2.5 ^b	14.4 ± 8.2	11.6 ± 3.6			
12	12.7 ± 11.1 ^b	21.8 ± 4.7 ^c	21.8 ± 5.4 ^b	160.7 ± 26.2 ^b	13.3 ± 6.7 ^b	3.9 ± 1.9 ^b	21.4 ± 2.6 ^b	14.2 ± 11.2	11.1 ± 3.9 ^c			
13	15.8 ± 11.8 ^b	23.2 ± 4.7 ^b	26.0 ± 7.0 ^b	174.6 ± 29.8 ^b	14.1 ± 7.0 ^b	4.5 ± 2.2 ^b	20.5 ± 2.5 ^b	12.6 ± 2.3	10.4 ± 4.0			
14	20.0 ± 14.0 ^b	24.1 ± 4.4 ^b	31.3 ± 8.3 ^b	189.7 ± 33.3 ^b	16.1 ± 8.1 ^b	5.3 ± 2.0 ^b	19.9 ± 2.4 ^b	12.2 ± 6.7	10.1 ± 3.7			
15	25.0 ± 15.2 ^b	25.1 ± 4.4 ^b	36.4 ± 7.6 ^b	200.0 ± 35.2 ^b	17.1 ± 8.1 ^b	5.5 ± 2.1 ^b	19.5 ± 2.5 ^b	11.4 ± 2.2	8.8 ± 4.2			
16	28.2 ± 15.7 ^b	25.5 ± 4.7 ^c	40.5 ± 7.5 ^b	212.4 ± 34.1 ^b	19.2 ± 9.4 ^b	6.5 ± 2.3 ^b	19.1 ± 3.1 ^b	10.9 ± 2.5	9.2 ± 4.3			
17	30.8 ± 15.4 ^b	25.8 ± 4.5 ^b	41.0 ± 7.5 ^b	220.5 ± 28.7 ^b	20.0 ± 9.4 ^b	5.8 ± 2.4 ^b	18.7 ± 2.6 ^b	11.3 ± 3.0	9.2 ± 4.3			
Girls (<i>n</i> = 4957)												
6	5.6 ± 6.2	11.8 ± 5.2	8.0 ± 2.4	100.1 ± 19.7	18.3 ± 5.1	–	26.7 ± 2.8	22.8 ± 4.3	14.5 ± 3.0			
7	5.6 ± 6.3	14.3 ± 4.7	9.4 ± 3.0	108.3 ± 23.5	17.8 ± 5.7	1.6 ± 0.8	25.5 ± 2.7	20.3 ± 3.7	13.4 ± 2.5			
8	6.5 ± 6.8	15.6 ± 4.8	11.4 ± 3.1	121.6 ± 19.7	17.4 ± 6.5	1.8 ± 0.9	24.4 ± 2.6	18.0 ± 3.0	13.1 ± 2.9			
9	6.6 ± 6.5	17.0 ± 4.7	12.8 ± 3.4	129.3 ± 22.4	16.5 ± 6.4	1.9 ± 0.9	23.9 ± 2.5	16.2 ± 2.8	12.4 ± 3.2			
10	7.8 ± 7.4	18.4 ± 4.6	14.8 ± 3.9	135.7 ± 25.5	16.3 ± 6.7	2.6 ± 1.3	22.9 ± 2.5	14.9 ± 4.7	11.9 ± 3.5			
11	7.6 ± 7.6	18.7 ± 4.4	17.2 ± 4.8	142.7 ± 23.1	16.6 ± 6.9	2.7 ± 1.4	22.5 ± 2.4	14.2 ± 8.4	11.1 ± 3.8			
12	7.6 ± 7.6	19.1 ± 4.3	19.9 ± 5.5	147.4 ± 25.5	17.4 ± 7.1	2.9 ± 1.5	22.3 ± 2.5	13.3 ± 2.6	11.1 ± 5.1			
13	8.2 ± 8.8	20.3 ± 4.5	23.1 ± 5.3	154.1 ± 27.6	19.5 ± 8.3	3.3 ± 1.7	21.8 ± 2.6	12.7 ± 2.3	10.1 ± 3.8			
14	9.1 ± 9.1	20.0 ± 4.4	26.1 ± 5.6	159.8 ± 30.3	21.0 ± 8.0	3.3 ± 1.5	21.2 ± 2.2	12.7 ± 7.8	9.7 ± 4.1			
15	9.6 ± 8.9	20.6 ± 4.9	27.0 ± 5.1	161.7 ± 26.3	22.2 ± 8.3	3.3 ± 1.6	21.2 ± 2.1	11.9 ± 2.6	8.7 ± 4.5			
16	11.5 ± 11.1	20.9 ± 4.6	27.8 ± 5.5	164.4 ± 26.9	23.5 ± 9.5	3.9 ± 1.4	21.0 ± 2.4	12.0 ± 9.3	8.4 ± 4.2			
17	11.6 ± 11.0	21.9 ± 4.3	28.5 ± 6.2	170.0 ± 23.6	24.2 ± 8.3	3.6 ± 1.4	20.6 ± 2.4	12.1 ± 8.1	8.4 ± 4.2			

^aBoys, $n = 2348$; girls, $n = 2011$.^bStatistical differences between the sexes at the level $p \leq 0.001$.^cStatistical differences between the sexes at the level $p \leq 0.05$.

Table II. Normal values (in repeats) of the physical fitness of Latvian school-age children: trunk strength – sit-ups (SUP) test.

Age (years)	Percentiles										
	5	10	20	30	40	50	60	70	80	90	95
Boys (<i>n</i> = 5507)											
6	2.0	5.0	7.0	10.0	11.0	12.0	13.0	15.0	17.0	20.0	21.0
7	7.0	9.0	12.0	13.0	14.0	15.0	17.0	18.0	19.0	20.0	22.0
8	8.0	11.0	14.0	15.0	16.0	17.0	18.0	20.0	21.0	23.0	24.0
9	10.0	13.0	15.0	17.0	18.0	19.0	20.0	21.0	23.0	24.0	26.0
10	12.0	14.0	17.0	18.0	19.0	20.0	21.0	23.0	24.0	26.0	27.0
11	14.0	16.0	18.0	19.0	20.0	21.0	22.0	24.0	25.0	27.0	28.0
12	15.0	16.0	18.0	20.0	21.0	22.0	23.0	24.0	25.0	28.0	29.0
13	16.0	18.0	20.0	21.0	22.0	23.0	24.0	26.0	27.0	29.0	30.0
14	17.0	19.0	20.0	22.0	23.0	24.0	25.0	26.0	28.0	30.0	32.0
15	18.0	20.0	21.0	22.0	24.0	25.0	26.0	28.0	29.0	30.0	32.0
16	18.0	20.0	22.0	23.0	25.0	26.0	27.0	28.0	29.0	31.0	32.0
17	19.0	20.0	22.0	24.0	25.0	26.0	27.0	28.0	30.0	31.0	32.0
Girls (<i>n</i> = 4957)											
6	1.0	4.0	7.0	10.0	11.0	12.0	13.0	14.0	15.0	18.0	20.0
7	6.0	8.0	11.0	12.0	14.0	15.0	16.0	17.0	18.0	20.0	21.0
8	7.0	9.0	12.0	14.0	15.0	16.0	17.0	18.0	19.0	21.0	23.0
9	8.0	11.0	14.0	15.0	16.0	17.0	18.0	19.0	20.0	23.0	24.0
10	11.0	13.0	15.0	16.0	18.0	19.0	20.0	21.0	22.0	24.0	25.0
11	11.0	13.0	15.0	17.0	18.0	19.0	20.0	21.0	22.0	24.0	26.0
12	12.0	13.0	16.0	17.0	19.0	20.0	20.0	22.0	23.0	24.0	26.0
13	13.0	15.0	17.0	18.0	19.0	20.0	21.0	22.0	24.0	26.0	28.0
14	14.0	15.0	17.0	18.0	19.0	20.0	21.0	22.0	23.0	25.0	27.0
15	13.0	15.0	17.0	18.0	20.0	21.0	22.0	23.0	25.0	26.0	29.0
16	13.0	15.0	17.0	19.0	20.0	21.0	22.0	23.0	25.0	27.0	28.0
17	14.0	18.0	19.0	20.0	21.0	22.0	23.0	24.0	25.0	27.0	28.0

Table III. Normal values (in min) of the physical fitness of Latvian school-age children: cardiorespiratory endurance – endurance shuttle run (ESR) test.

Age (years)	Percentiles										
	5	10	20	30	40	50	60	70	80	90	95
Boys (<i>n</i> = 2348)											
7	0.5	1.1	1.1	1.2	1.3	1.5	2.0	2.3	2.6	3.2	4.0
8	1.0	1.1	1.2	1.3	1.5	2.0	2.3	3.0	3.3	4.1	4.5
9	0.6	1.1	1.3	1.4	2.1	2.3	2.5	3.2	4.0	5.0	5.3
10	1.0	1.2	1.4	2.1	2.4	3.1	3.5	4.1	5.0	5.5	6.4
11	1.1	1.4	2.0	2.3	3.0	3.4	4.1	4.6	5.3	6.4	7.1
12	1.3	1.5	2.2	2.5	3.2	3.4	4.3	5.1	5.6	6.5	7.8
13	1.2	1.4	2.3	3.2	3.6	4.4	5.1	5.5	6.4	7.4	8.2
14	2.3	3.0	3.4	4.2	4.5	5.2	5.5	6.3	7.1	8.1	9.1
15	2.4	3.1	3.5	4.1	4.6	5.3	6.1	6.5	7.3	8.2	9.3
16	2.5	3.2	4.4	5.4	6.1	6.4	7.2	7.5	8.4	10.0	10.5
17	2.3	3.1	4.1	4.5	5.2	6.1	6.4	7.2	7.5	8.2	9.5
Girls (<i>n</i> = 2011)											
7	0.5	1.0	1.1	1.2	1.3	1.4	1.5	2.0	2.2	3.0	3.2
8	0.9	1.0	1.1	1.2	1.3	1.4	2.0	2.2	2.4	3.3	3.5
9	1.0	1.1	1.2	1.3	1.4	1.5	2.1	2.3	2.6	3.3	4.1
10	1.0	1.2	1.3	1.5	2.1	2.3	2.6	3.2	4.0	4.5	5.2
11	1.1	1.2	1.4	1.7	2.2	2.4	3.0	3.3	4.1	4.9	5.3
12	1.1	1.2	1.4	2.1	2.2	2.5	3.2	3.5	4.2	5.1	5.5
13	1.2	1.3	1.6	2.2	2.5	3.2	3.5	4.2	4.5	5.9	6.3
14	1.3	1.5	2.1	2.4	2.6	3.3	3.4	4.2	4.5	5.3	6.1
15	1.1	1.4	1.5	2.2	2.4	3.1	3.3	4.1	5.0	5.3	6.2
16	1.5	2.1	2.4	3.1	3.3	3.5	4.1	4.5	5.2	6.3	7.1
17	1.8	2.2	2.3	2.5	3.2	3.3	4.1	4.3	5.0	5.2	6.6

Table IV. Normal values (in s) of the physical fitness of Latvian school-age children: speed of movement/agility – shuttle run test 10×5 m (SHR).

Age (years)	Percentiles										
	5	10	20	30	40	50	60	70	80	90	95
Boys ($n = 5507$)											
6	21.7	22.8	23.7	24.3	25.1	25.7	26.4	27.1	28.5	30.1	31.4
7	20.5	21.2	22.3	23.1	23.7	24.4	25.0	25.8	27.0	28.2	29.1
8	19.8	20.4	21.5	22.1	22.8	23.5	24.2	24.9	25.8	27.3	29.1
9	19.4	20.0	20.8	21.5	22.1	22.8	23.2	24.0	24.8	26.3	27.8
10	18.6	19.4	20.2	20.7	21.3	22.1	22.7	23.2	24.0	25.5	26.8
11	18.4	19.0	19.7	20.4	21.1	21.5	22.1	22.6	23.8	25.1	26.5
12	17.5	18.2	19.3	20.0	20.7	21.2	21.8	22.5	23.5	24.7	26.4
13	16.8	17.7	18.3	19.0	19.7	20.3	20.9	21.5	22.3	23.7	25.0
14	16.1	16.8	17.9	18.7	19.3	20.0	20.5	21.2	21.9	22.9	24.0
15	15.8	16.5	17.7	18.4	18.8	19.4	19.8	20.3	21.1	22.6	23.8
16	15.2	15.6	16.7	17.6	18.1	18.8	19.4	20.0	20.8	22.7	24.4
17	15.2	15.5	16.4	17.2	18.1	18.7	19.1	19.7	20.4	21.6	23.8
Girls ($n = 4957$)											
6	22.4	23.4	24.5	25.2	25.6	26.4	27.1	27.9	28.6	30.6	32.1
7	21.2	22.8	23.9	24.6	25.3	26.1	26.9	27.8	29.4	29.3	30.5
8	20.5	21.3	22.3	23.1	23.7	24.3	24.9	25.5	26.5	27.9	28.9
9	20.3	21.0	21.9	22.3	23.0	23.6	24.4	25.2	25.9	27.1	28.6
10	19.2	20.0	21.2	21.8	22.3	22.8	23.4	24.0	24.8	25.9	27.5
11	18.6	19.6	20.6	21.3	21.9	22.4	23.0	23.6	24.3	25.4	26.7
12	18.6	19.8	20.4	21.1	21.7	22.2	22.8	23.4	24.1	25.2	26.2
13	18.2	19.1	20.0	20.5	21.0	21.5	22.3	22.8	23.5	24.8	25.9
14	17.7	18.4	19.4	19.9	20.5	21.1	21.6	22.2	22.9	24.2	25.0
15	18.0	18.9	19.6	20.1	20.6	21.6	21.6	22.1	22.9	24.0	25.2
16	17.1	18.2	19.0	19.9	20.6	21.1	21.5	22.1	22.6	23.9	24.9
17	17.1	17.9	18.6	19.4	19.9	20.6	21.0	21.7	22.2	23.2	24.0

considerable differences. The mean SBJ test score difference between our study and the HELENA study was 17.5 cm for boys and 21.8 cm for girls. Comparing our study with the HELENA study for the SAR test, the mean score difference was -5.2 cm for boys and -10.9 cm for girls. The mean score difference for the cardiorespiratory endurance test (Figure 1e) was -0.8 min for boys and -0.16 min for girls. The running distance used for the speed test differed between the studies, but the trends in the performances were similar (Figure 1f). The values from our study should therefore be appropriate for application by those practitioners and researchers who implement the 10×5 m version of the SHR test instead of the 10×4 m version. These comparative observations imply that the transfer of reference values for physical fitness between populations and geographic regions is not straightforward, and suggest that specific norms need to be developed for each population in which testing is to be performed.

Comparing the current data with other studies involving homogeneous populations of European youth, we find that adolescents from Spain [16]

performed better on the cardiorespiratory endurance test than the Latvian adolescents. The authors from Spain also analysed the results of 15 studies from different countries and concluded that the performance of Spanish adolescents was worse than the performances reported from 11 of the 15 studies. Our study population thus showed a lower level of cardiorespiratory endurance compared with homogeneous populations of adolescents investigated in previous studies. This can be explained by an insufficient infrastructure for participation in sports and spontaneous physical activities, and a lack of policies for promotion of habitual physical activity among school-aged children. Children from lower social classes are not always able to participate in organised sports activities because of lack of financial resources. One similarity to a Swiss study performed in 1996–1997 [20] is that the performance among boys improved with age for all tests except for the flexibility test, whereas improvement among girls decreased after the onset of menstruation. Comparing the mean SAR test scores from Latvian and Swiss [20] pupils between 9 and 17 years of age, and between Latvia

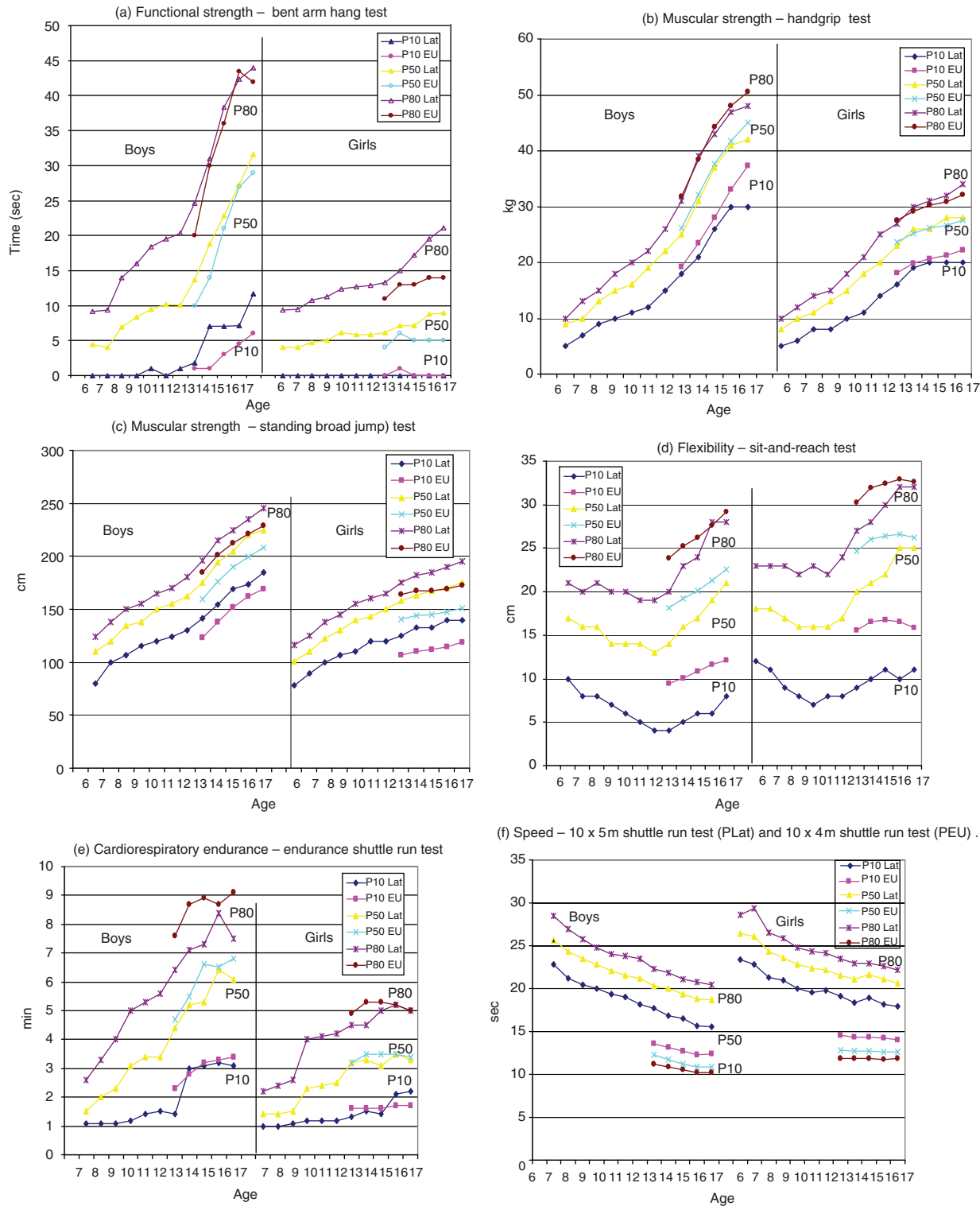


Figure 1. Physical fitness percentile curves (P₁₀, P₅₀, P₈₀) of six EUROFIT tests from the present study (PLat) and from 10 European countries (PEU) [19]. (a) Functional strength – bent arm hang test; (b) Muscular strength – handgrip test; (c) Muscular strength – standing broad jump test; (d) Flexibility – sit-and-reach test; (e) Cardiorespiratory endurance – endurance shuttle run test; (f) Speed – 10x5m shuttle run test (PLat) and 10x4m shuttle run test (PEU).

and Spanish pupils [21] between 13 and 17 years of age, the adolescents from Switzerland scored better in all age groups and both genders. Compared with mean scores from the trunk strength (SUP) test for 13-year-old boys reported from Greece [22], Latvian boys showed worse results. Comparing our data on muscular strength with data from Lithuania collected in 2002, the Latvian adolescents aged 12, 14, and 16 years showed slightly worse results among both boys and girls [23]. Even though these observations may seem discouraging for Latvian adolescents, it must be remembered that our study data were collected from a common population of school-aged children and adolescents with a low level of non-participation in the study (5%). More studies of physical fitness involving representative populations of children and adolescents are therefore warranted before conclusions can be drawn about country-level differences.

Conclusions

Planning of evidence-based programs for promotion of physical activity among the young requires reference values for physical fitness in different age groups. We have established standardised values for health-related physical fitness among northern European youth based on data from Latvian children and adolescents aged between 6 and 17 years. Bearing in mind that the objective for physical education is not physical fitness *per se*, the reference values can be used as a part of individual-level assessments of students. Maybe even more importantly, they allow monitoring of population-based trends at regional and national levels. More research is needed to establish validated cut-off values for pathological cardiorespiratory endurance and muscular strength tests.

Declaration on conflict of interest

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgements

The study was financed by regular funding to the Sports Medicine State Agency from the Latvian Ministry of Health. Reference value tables for all tests are available upon request from the corresponding author.

References

- [1] Bouchard C, Sheppard RJ. Physical activity, fitness, and health: the model and key concepts. In: Bouchard C, Sheppard RJ, Stephens T, editors. Physical activity, fitness, and health. Champaign, IL: Human Kinetics Books; 1994. pp. 77–88.
- [2] CDC/US Department of Health and Human Services. Physical activity and health: a report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centres for Disease Control and Prevention, National Centre for Chronic Disease Prevention and Health Promotion, <http://www.cdc.gov/nccdphp/sgr/sgr.htm> (accessed September 2009).
- [3] Pate RR. A new definition of youth fitness. *Phys Sports Med* 1983;11:77–83.
- [4] Suni J, Oja P, Laukkanen R, Mäilänpalo SI, Pasanen ME, Vuori IM, et al. Health-related fitness test battery for adults: aspects of reliability. *Arch Phys Med Rehabil* 1996;77(4):399–405.
- [5] Ruiz JR, Ortega FB, Gutierrez A, Meusel D, Sjöström M, Castillo MJ. Health-related fitness assessment in childhood and adolescence: a European approach based on the AVENA, EYHS and HELENA studies. *J Public Health* 2006;14:269–77.
- [6] Ruiz JR, Rizzo N, Wennlof A, Ortega FB, Harro M, Sjöström M. Relations of total physical activity and intensity to fitness and fatness in children: the European Youth Heart Study. *Am J Clin Nutr* 2006;84(2):299–303.
- [7] Bergström E, Hernell O, Persson A. Endurance running performance in relation to cardiovascular risk indicators in adolescents. *Int J Sports Med* 1997;18(4):300–7.
- [8] Nielsen GA, Andersen LB. The association between high blood pressure, physical fitness, and body mass index in adolescents. *Prev Med* 2003;36(2):229–34.
- [9] Hasselström H, Hansen SE, Froberg K, Andersen LB. Physical fitness, physical activity during adolescence as predictors of cardiovascular disease risk in young adulthood. Danish Youth and Sports Study. An eight-year follow-up study. *Int J Sports Med* 2002;23(Suppl 1):27–31.
- [10] Ruiz JR, Castro-Piñero J, Artero EG, Ortega FB, Sjöström M, Suni J, et al. Predictive validity of health-related fitness in youth: a systematic review. *Br J Sports Med* 2009;43(12):909–23.
- [11] Vicente-Rodriguez G, Ara I, Perez-Gomez J, Serrano-Sanchez JA, Dorado C, Calbet JA. High femoral bone mineral density accretion in prepubertal soccer players. *Med Sci Sports Exerc* 2004;36(10):1789–95.
- [12] Vicente-Rodriguez G, Ara I, Perez-Gomez J, Dorado C, Calbet J. Muscular development and physical activity as major determinants of femoral bone mass acquisition during growth. *Br J Sports Med* 2005;39(9):611–16.
- [13] Oja P, Tuxworth B. Eurofit for adults. Assessment of health-related fitness. Tampere, Finland: Council of Europe; 1995.
- [14] Vanhees L, Lefevre J, Philippaerts R, Martens M, Huygens W, Troosters T, et al. How to assess physical activity? How to assess physical fitness. *Eur J Cardiovasc Prev Rehabil* 2005;12(2):102–14.
- [15] Committee of Experts on Sports Research. EUROFIT: Handbook for the EUROFIT tests of physical fitness, 2nd edn. Strasbourg: Committee of Experts on Sports Research; 1993.
- [16] Ortega FB, Artero EG, Ruiz JR, Vicente-Rodriguez G, Bergman P, Hagströmer M, et al. Reliability of health-related physical fitness tests in European adolescents.

- The HELENA Study. *Int J Obes (Lond)* 2008;32 (Suppl 5):S49–57.
- [17] Castro-Piñero J, Artero EG, España-Romero V, Ortega FB, Sjöström M, Suni J, et al. Criterion-related validity of field-based fitness tests in youth: a systematic review. *Br J Sports Med* 2009;27:405–13.
- [18] Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: a powerful marker of health in young people. *Int J Obesity* 2008;32(1):1–11.
- [19] Ortega FB, Artero EG, Ruiz JR, España-Romero V, Jiménez-Pavón D, Vicente-Rodríguez G, et al. Physical fitness levels among European adolescents: the HELENA study. *Br J Sports Med* 2009; doi:10.1136/bjsm.2009.062679.
- [20] Narring F, Cauderay M, Cavadini C, Michaud PA. Physical fitness and sport activity of children and adolescents: methodological aspects of a regional survey. *Soz- Präventivmed* 1999;44(2):44–54.
- [21] Ortega FB, Ruiz JR, Castillo MJ, Moreno LA, Urzanqui A, González-Gross M, et al. Health-related physical fitness according to chronological and biological age in adolescents. The AVENA study. *J Sports Med Phys Fitness* 2008;48(3):371–9.
- [22] Koutedakis Y, Bouziotas C. National physical education curriculum: motor and cardio vascular health related fitness in Greek adolescents. *Br J Sports Med* 2003;37(4):311–14.
- [23] Volbekiene V, Gričiute A. Health-related physical fitness among schoolchildren in Lithuania: a comparison from 1992 to 2002. *Scand J Public Health* 2007;35(3):235–42.

ORIGINAL ARTICLE

Physical fitness normative values for 6–18-year-old Greek boys and girls, using the empirical distribution and the lambda, mu, and sigma statistical method

KONSTANTINOS D. TAMBALIS^{1,2}, DEMOSTHENES B. PANAGIOTAKOS¹,
GLYKERIA PSARRA¹, STELIOS DASKALAKIS³, STAVROS A. KAVOURAS^{1,4},
NICKOS GELADAS², SAVAS TOKMAKIDIS⁵, & LABROS S. SIDOSSIS^{1,6}

¹Department of Nutrition and Dietetics, Harokopio University, Athens, Greece, ²Department of Physical Education and Sport Science, University of Athens, Athens, Greece, ³Ministry of Education and Religions, Athens, Greece, ⁴Department of Health, Human Performance, and Recreation, Human Performance Laboratory, University of Arkansas, Fayetteville, AR, USA, ⁵Department of Physical Education and Sport Science, Dimokritio University, Komotini, Greece, ⁶Department of Internal Medicine, Shriners Hospital for Children, University of Texas Medical Branch, Galveston, TX, USA

Abstract

The aim of this study was to establish age- and gender-specific physical fitness normative values and to compare percentiles and Z scores values in a large, nationwide sample of Greek children aged 6–18 years. From March 2014 to May 2014, a total of 424,328 boys and girls aged 6–18 years who attended school in Greece were enrolled. The studied sample was representative, in terms of age–sex distribution and geographical region. Physical fitness tests (i.e. 20 m shuttle run test (SRT), standing long jump, sit and reach, sit-ups, and 10 × 5 m SRT) were performed and used to calculate normative values, using the percentiles of the empirical distributions and the lambda, mu, and sigma statistical method. Normative values were presented as tabulated percentiles for five health-related fitness tests based on a large data set comprising 424,328 test performances. Boys typically scored higher than girls on cardiovascular endurance, muscular strength, muscular endurance, and speed/agility, but lower on flexibility (all *p* values < 0.001). Older boys and girls had better performances than younger ones (*p* < 0.001). Physical fitness tests' performances tended to peak at around the age of 15 years in both sexes. The presented population-based data are the most up-to-date sex- and age-values for the health-related fitness of children and adolescents in Greece and can be used as standard values for fitness screening and surveillance systems and for comparisons among the same health-related fitness scores of children from other countries similar to Greece. Schools need to make efforts to improve the fitness level of the schoolchildren through the physical education curriculum to prevent cardiovascular risk.

Keywords: *Physical fitness, normative values, performance, children*

Introduction

Physical fitness refers to “the ability to carry out daily tasks with vigor and alertness, without undue fatigue and with ample energy to enjoy leisure-time pursuits and to meet unforeseen emergencies” (Caspersen, Powell, & Christenson, 1985, p. 128). Fitness status has long been associated with age, from youth to the middle years of lifespan and even older. However, there is a lack of data in teenagers and younger children. It is crucial to know the

fitness level of children as it has been suggested that fitness level in childhood is essential to carry forward favourable behavioural and biological effects into later life (Ortega, Ruiz, Castillo, & Sjöström, 2008). Accumulating epidemiologic evidence reveals that improvement in physical fitness, mainly aerobic fitness, is related to better health in children (Andersen, Wedderkopp, Hansen, Cooper, & Froberg, 2003; Ekelund et al., 2007; Hurtig-Wennlof, Ruiz, Harro, & Sjöström, 2007; Tambalis,

Correspondence: L. S. Sidossis, Department of Internal Medicine, Shriners Hospital for Children, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555–0177, USA, E-mail: lasidoss@utmb.edu

Panagiotakos, Psarra, & Sidossis, 2011), in a dose response manner (Anderssen et al., 2007). Moreover, subjects with high physical fitness during adolescence may have lower levels of body fatness as adults (Eisenmann, Wickel, Welk, & Blair, 2005). In contrast, low levels of physical fitness in children have been associated with a number of cardiometabolic risk factors, such as hypertension, hyperlipidaemia, and obesity (Anderssen et al., 2007). To prevent early development of cardiovascular disease risk factors, among other significant interventions (e.g. increased physical activity levels, decreased obesity levels), preventive strategies must incorporate age- and sex-specific physical fitness assessment, even from childhood.

To make useful recommendations as regards to physical fitness, percentiles and *Z* scores have been used to assess children's and adolescent's growth and fitness status (Eisenmann, Laurson, & Welk, 2011). Percentiles are easier to understand and utilize in practice. In contrast, *Z* scores are more complicated and may be of limited use in clinical settings, but are more useful in research. However, it has been suggested that the lambda, mu, and sigma (LMS) method, a technique that uses *Z* scores and is based on smoothed percentile distribution curves over ages, better fits the data than the empirical percentile method (Wang, Moreno, Caballero, & Cole, 2006). These different methodologies may affect the cut-off points for the evaluation of fitness status and may lead to different results.

Numerous data sets on physical fitness levels among children are available worldwide; moreover age-, gender-, as well as region-specific values are essential for a variety of health-risk measurements (i.e. not only for fitness), in order to develop effective public health strategies (Tambalis et al., 2011; Tambalis, Panagiotakos, Arnaoutis, & Sidossis, 2013; Tokmakidis, Kasambalis, & Christodoulos, 2006). In addition, a comparison of both methodologies may prove useful for future researchers who would also like to calculate country-specific normative curves for fitness levels, as it would allow them to select the most appropriate procedure.

Therefore, the aim of the present work was to present the Greek region-specific distribution of age and sex physical fitness test measurements for 6–18-year-old children and adolescents and to evaluate sex- and age-related differences, using both percentiles and *Z* scores values in a nationwide sample of schoolchildren. These distributions (percentiles and *Z* scores) may also be valid for other similar populations around the world for comparisons among the same health-related fitness scores of children similar to Greece (i.e. Caucasians), and guide

preventive strategies to better prevent cardiometabolic disorders in the future.

Methods

Study design and participants

Population-based, representative data were derived from a nationwide school-based survey under the auspices of the Ministry of Education. Specifically, anthropometric, physical activity, nutrition, and physical fitness data along with information on age and sex were collected from March 2014 to May 2014. In total, 424,328 (51% boys and 49% girls) children aged 6–18 years from Greek public and private schools agreed to participate in the study (participation rate was 40% of the total population). The working sample was representative of the entire Greek population (chi-square *p* value as compared with the current sample with the age–sex distribution of all Greek areas = 0.93).

Study approval

Ethical approval for the survey was granted by the Review Board of the Ministry of Education and the Ethical Committee of Harokopio University.

Assessment of fitness status

The Euro-fit physical fitness test battery was used to evaluate children's physical fitness levels; it is a set of nine physical tests covering flexibility, speed, endurance, and strength. The standardized test battery was proposed by the Council of Europe for children of school age and has been used in many European countries schools since the 1980s. Five fitness tests of the Euro-fit Battery, representative of flexibility, explosive strength, speed/agility, and aerobic performance, were administered by trained physical education professionals during the physical education classes. Specifically, measurements were performed by one teacher of physical education in each class. All physical education professionals were instructed through a detailed and extended manual of operations and followed a standardized procedure of measurements in order to minimize the potential inter-rate variability among schools. The physical education teachers were first trained by a school advisor of physical education for accurate anatomical landmarks, subject positioning, and measurement techniques. Verbal informed consent for the child to participate in the measurements was taken from physical education teachers. As the measurements were included in an obligatory school programme, verbal informed consent was considered sufficient.

Briefly, the test assessed: (a) standing long jump (SLJ; jump as far as possible from a standing position at the start) to evaluate lower body explosive power; (b) sit and reach (SR; this test involves sitting on the floor with legs stretched out straight ahead without shoes. The soles of the feet are placed flat against the box. With the palms facing downwards, and the hands on top of each other or side by side, the participant reaches forward along the measuring line as far as possible. The score is recorded to the nearest centimetre as the distance reached by the hand, using 15 cm at the level of the feet) to measure flexibility; (c) sit-ups in 30 s (SUs; lie on the mat with the knees bent at right angles, with the feet flat on the floor and held down by a partner), to measure the endurance of the abdominal and hip-flexor muscles; (d) 10 × 5 m shuttle run test (10 × 5 m SRT; from a standing start), to evaluate speed and agility; and (e) multi-stage 20 m SRT, to estimate aerobic performance. The 20 m SRT test consists of measuring the number of laps completed by participants running up and down between two lines in groups, set 20 m apart, at an initial speed of 8.5 km/h which increases by 0.5 km/h every minute, using a pre-recorded audio tape (Leger, Lambert, Goulet, Rowan, & Dinelle, 1984; Leger, Mercier, Gadoury, & Lambert, 1988). Repeat tests (two trials) were allowed for the SLJ, SR, SU, and 10 × 5 m SRT, with the best performance of each recorded.

Statistical analysis

Descriptive statistics (mean ± standard deviation) for boys and girls were calculated. Comparisons of the physical fitness tests' performances data between boys and girls were performed using the independent samples *t* test, after testing for equality of variances using the Levene test. Comparisons between percentile values of the physical fitness tests' performance data of both calculation methods were performed using the paired samples *t* test.

Age- and sex-specific distributions and percentiles were calculated using two methods: through the empirical distribution of the data, the 3th, 10th, 25th, 50th, 75th, 90th, and 97th percentiles were calculated; also, the LMS method was used (Cole & Green, 1992). The LMS method was used in order to smooth the age-dependent skewness usually observed in fitness values. Based on this method, the data were normalized using the Box-Cox power transformation. The Box-Cox λ -power transformation of the variable y_i has the following form:

$$y_i^{(\lambda)} = \frac{y_i^\lambda - 1}{\lambda(\text{gm}(y))^\lambda - 1} \quad \text{if } \lambda \neq 0$$

or

$$y_i^{(\lambda)} = \text{gm}(y_i) \log y_i \quad \text{if } \lambda \neq 0$$

where gm is the geometric mean of y_i . The power transformation was calculated from the raw data and skewness in the distribution where y was removed. The principle idea of the LMS method is to power transform the measurement, i.e. SLJ here, and to use the coefficient of variation (CV = standard deviation/mean) of the raw data. The optimal Box-Cox power λ is the one that gives the lowest CV (Cole & Green, 1992). Thus, the LMS method calculates the best power (L), the best mean (M), and CV (S) in each series of measurements at a specific age. The degrees of freedom used to determine the L , M , and S were chosen on the basis of that which achieved the smallest difference in the penalized deviance ($-2 \times \log\{\text{penalised likelihood}\}$) statistic, as well as the Schwarz Bayesian Criterion (both goodness-of-fit measures) between the estimated models. Then, centiles of physical fitness tests were calculated as follows:

$$\text{Centile}(\alpha) = M_{\text{age}} \times (1 + L_{\text{age}} \times S_{\text{age}} \times Z_{\alpha})^{1/L_{\text{age}}}, \quad (1)$$

where Z_{α} is the Z score (i.e. (variable – mean)/standard deviation) corresponding to the required centile (e.g. $Z = -0.67$ gives the 25th centile, $Z = 0$ gives the median M , $Z = 0.67$ gives the 75th centile). The gender-specific cut-offs of fitness tests for each age group were calculated for various centiles by solving the previous equation. Particularly, from Equation (1) we have:

$$\begin{aligned} \left(\frac{\text{Centile}(\alpha)}{M_{\text{age}}} \right)^{L_{\text{age}}} &= 1 + L_{\text{age}} \times S_{\text{age}} \times Z_{\alpha}, \\ \left(\frac{\text{Centile}(\alpha)}{M_{\text{age}}} \right)^{L_{\text{age}}} - 1 &= L_{\text{age}} \times S_{\text{age}} \times Z_{\alpha}. \end{aligned}$$

Thus,

$$Z_{\alpha} = \left(\left(\frac{\text{Centile}(\alpha)}{M_{\text{age}}} \right)^{L_{\text{age}}} - 1 \right) \times (L_{\text{age}} \times S_{\text{age}})^{-1}. \quad (2)$$

The LMS values can be used to calculate Z scores and therefore percentile values by looking up a Z

table, using the following formula:

$$z = \frac{(x/M) - 1}{L \times S},$$

where x is performance, L is the gender- and age-specific L value, M is the gender- and age-specific M value, and S is the gender- and age-specific S value. All statistical analyses were performed using the SPSS program (Release 18; SPSS Inc., Chicago, IL, USA). The LMSchartmaker (Pan & Cole, 2010) and the LMSgrowth (Pan & Cole, 2011) freeware packages were used to calculate L , M , and S values at ages 6–18 based on Greek reference values.

Results

In Tables I and II, normative physical fitness data for 6–18-year-old children in Greece, by gender and age as tabulated critical percentiles and LMS values from 3 to 97 (P_3 , P_{10} , P_{25} , P_{50} , P_{75} , P_{90} , P_{97}), are presented. Also presented are the gender- and age-specific LMS values for all fitness tests.

For each of the fitness tests, performance was better in boys compared with girls ($p < 0.001$), except for the SR test ($p < 0.001$). Moreover, older boys and girls had better performances than younger ones ($p < 0.001$). Physical fitness tests' performances also tended to peak at about the age of 15–16 years in both sexes.

In order to investigate potential differences between percentile values from the two methods, comparisons by physical fitness test and gender were performed. Data analysis revealed no significant differences between critical percentiles and LMS percentiles in SR, SLJ, SUs, 20 m SRT, and 10 × 5 m SRT values in either sex (all p values > 0.05).

Discussion

The aim of the present work was to develop up-to-date age- and sex-specific physical fitness normative values for Greek children aged 6–18 years and to compare specific percentile values from two widely applied estimation methods: the frequency percentiles and the LMS smoothed percentiles. This study provides information relating to normative values across a range of health-related fitness tests. These values could be used as approximate indicative values for comparisons among the same health-related fitness scores of children (according to National Statistical Services the vast majority of children population in Greece are Caucasians) from

other countries similar to Greece: i.e. a developed country with population predominately Caucasian. Moreover, these data can be used as benchmark values for health-related fitness screening and surveillance of children 6–18 years in Greece.

The presented data could be useful in a crude classification of how well children in Greece perform on the specific health-related fitness tests relative to their age- and sex-matched peers and to identify those children with specific physical fitness characteristics that could be considered essential for sporting success. Moreover, our findings facilitate the recognition of children and adolescents with low physical fitness aiming to set appropriate goals in the future and to promote encouraging health behaviours. Previous studies in children from Greece have linked low cardiorespiratory fitness with increased metabolic score risk and inflammation (Christodoulos, Douda, & Tokmakidis, 2012; Flouris, Bouziotas, Christodoulos, & Koutedakis, 2008). According to the presented percentile classification from both methods, and for a practical use of these data, children could be classified as having a performance score: very poor ($< P_{10}$), in the poor quartile (1st), in the good quartiles (2nd–3rd), in the very good quartile (4th), and excellent ($X > P_{90}$) of the distribution. Although this classification is not criterion-referenced, Looney and Plowman (1990) suggested that test scores above the 25th percentile (the poor quartile) in fitness tests should be considered acceptable from a health perspective. On the contrary, low levels of physical fitness have been associated with many serious health problems in childhood, while sufficient levels of physical fitness may have an important cardio-protective role in children. According to Anderssen et al. (2007), there was a strong association between aerobic fitness and the clustering of cardiovascular disease risk factors. Specifically, the odds ratios for clustering in each quartile of fitness, using the quartile with the highest fitness as reference, were 13.0 [95% confidence interval (CI) 8.8–19.1] 4.8 [95% CI 3.2–7.1], and 2.5 [95% CI 1.6–3.8], respectively, after adjusting for several confounding factors (Andersen et al., 2003; Anderssen et al., 2007; Ekelund et al., 2007; Hurtig-Wennlof et al., 2007). Future studies need to examine which specific childhood and/or adolescence thresholds for aerobic fitness are significantly associated with clustered cardiovascular disease risk factors.

The presented results revealed that boys consistently scored higher than girls in almost all fitness tests (with the exception of the SR test of flexibility). Moreover, older boys and girls performed better than younger ones ($p < 0.001$). Our findings are in accordance with recent studies from Latvia (Sauka et al.,

Table I. Boys physical fitness tests percentiles and LMS values and LMS summary statistics by age in 6–18-year-old Greek children and adolescents

Age	N	Percentiles							LMS method									
		P ₃	P ₁₀	P ₂₅	P ₅₀	P ₇₅	P ₉₀	P ₉₇	P ₃	P ₁₀	P ₂₅	P ₅₀	P ₇₅	P ₉₀	P ₉₇	L	M	S
20 m SRT (completed laps)																		
6	1706	1.0	4.0	9.0	14.0	22.0	31.0	44.2	2.2	4.9	9.0	14.6	22.1	31.6	43.2	0.38	14.64	0.67
7	15,196	5.0	8.0	12.0	18.0	28.0	38.0	50.0	3.2	6.7	11.8	18.9	28.0	39.4	53.2	0.39	18.86	0.64
8	20,774	6.0	9.0	14.0	23.0	34.0	47.0	59.0	4.2	8.6	14.8	23.1	33.6	46.6	62.1	0.41	23.06	0.61
9	20,149	7.0	11.0	17.0	28.0	41.0	53.0	64.0	5.2	10.4	17.6	27.1	38.9	53.0	69.7	0.44	27.09	0.59
10	19,279	8.0	12.0	20.0	31.0	45.0	58.0	70.0	6.0	12.1	20.5	31.1	44.1	59.3	76.8	0.49	31.13	0.57
11	18,793	9.0	15.0	23.0	36.0	51.0	64.0	78.0	6.6	13.8	23.4	35.4	49.5	65.8	84.1	0.54	35.36	0.55
12	11,456	8.0	15.0	24.0	39.0	54.0	69.0	82.0	6.8	15.4	26.4	39.7	55.1	72.4	91.6	0.60	39.71	0.54
13	15,555	7.0	16.0	28.0	44.0	61.0	76.0	91.0	6.8	16.7	29.3	44.0	60.7	79.0	98.9	0.66	44.04	0.54
14	13,681	7.0	19.0	32.0	50.0	69.0	85.0	100	6.4	17.8	31.8	47.8	65.4	84.5	105	0.73	47.79	0.53
15	10,135	9.0	20.0	34.0	53.0	72.0	90.0	103	5.5	18.3	33.5	50.4	68.6	87.9	108	0.79	50.37	0.52
16	1862	7.0	20.0	36.0	54.0	75.0	90.0	103	4.4	18.4	34.5	51.9	70.2	89.3	109	0.85	51.87	0.52
17	670	7.0	18.0	33.0	50.0	68.0	86.0	100	3.9	18.3	35.1	52.8	71.0	89.7	109	0.91	52.76	0.51
18	305	7.4	13.0	28.0	48.0	66.0	87.0	115	3.2	18.0	35.6	53.5	71.6	89.9	108	0.97	53.48	0.51
SLJ (cm)																		
6	17,494	64.0	76.0	90.0	102	115	128	140	59.0	75.0	89.7	103	116	129	141	1.42	103.4	0.19
7	23,994	71.0	85.0	100	113	127	140	151	66.0	83.1	98.7	113	127	141	153	1.42	113.4	0.19
8	23,836	80.0	94.0	108	123	138	150	162	72.8	90.7	107	123	137	151	165	1.43	122.7	0.18
9	22,667	85.0	100	116	131	147	160	172	79.0	97.7	115	131	146	161	175	1.44	131.2	0.18
10	21,886	90.0	108	123	140	155	169	181	84.8	104	122	139	155	171	185	1.44	139.2	0.18
11	21,241	97.0	113	130	147	163	177	190	90.3	111	130	147	164	180	196	1.45	147.3	0.18
12	12,961	100.0	119	135	154	171	187	201	95.9	117	137	156	174	191	207	1.45	156.0	0.17
13	8683	105.0	125	144	164	183	200	217	101	124	145	165	184	202	219	1.46	165.0	0.17
14	7231	112.0	133	153	175	195	212	228	106	130	152	173	193	212	230	1.48	173.3	0.17
15	2633	118	140	160	183	204	220	240	110	135	159	180	201	220	239	1.51	180.3	0.17
16	7729	122	145	167	188	209	225	240	112	139	163	186	207	226	245	1.55	185.6	0.17
17	2737	120	146	167	188	210	228	245	113	142	167	190	211	231	250	1.60	189.7	0.17
18	375	110	137	160	187	222	244	250	114	144	169	193	215	235	254	1.66	193.0	0.17
SR (cm)																		
6	21,448	2.0	5.0	10.0	15.0	18.0	22.0	25.0	2.1	5.7	9.6	13.7	18.0	22.4	27.0	0.87	13.70	0.46
7	27,509	2.0	5.0	9.0	15.0	18.0	22.0	25.0	1.8	5.3	9.2	13.4	17.9	22.5	27.3	0.84	13.44	0.48
8	27,220	2.0	4.0	9.0	14.0	18.0	22.0	26.0	1.6	5.0	8.9	13.1	17.7	22.5	27.5	0.81	13.14	0.51
9	26,692	2.0	4.0	8.0	14.0	18.0	22.0	25.5	1.4	4.6	8.5	12.8	17.4	22.4	27.6	0.78	12.78	0.53
10	25,017	2.0	4.0	8.0	13.0	17.0	21.5	26.0	1.3	4.3	8.1	12.4	17.1	22.2	27.5	0.76	12.44	0.54
11	24,418	1.0	3.0	7.0	13.0	17.0	21.0	26.0	1.2	4.2	8.0	12.3	17.1	22.2	27.7	0.74	12.29	0.56
12	15,636	1.0	3.0	7.0	13.0	17.0	22.0	27.0	1.2	4.2	8.0	12.5	17.4	22.8	28.5	0.72	12.47	0.57
13	14,613	1.0	3.0	8.0	14.0	18.0	23.0	28.0	1.3	4.4	8.4	13.0	18.2	23.8	29.9	0.71	12.99	0.57
14	12,871	2.0	4.0	9.0	15.0	20.0	25.0	30.0	1.4	4.7	8.9	13.8	19.3	25.3	31.7	0.71	13.80	0.57
15	9537	2.0	5.0	10.0	16.0	21.0	27.0	31.0	1.6	5.1	9.5	14.8	20.6	27.0	33.8	0.71	14.77	0.56
16	7291	2.0	5.0	11.0	16.5	22.0	28.0	32.0	1.8	5.5	10.2	15.8	21.9	28.7	35.9	0.70	15.76	0.56
17	2573	2.0	5.0	11.0	17.0	22.0	28.0	33.0	2.0	5.9	10.9	16.7	23.2	30.3	37.8	0.70	16.72	0.55
18	352	1.0	5.0	10.0	16.0	22.0	29.0	33.0	2.2	6.4	11.6	17.7	24.4	31.8	39.7	0.70	17.68	0.55

<i>SUs (no. in 30 sec)</i>																		
6	17,381	2.0	6.0	10.0	13.0	16.0	19.0	23.0	1.0	5.7	9.8	13.5	16.8	19.9	22.8	1.34	13.46	0.38
7	23,866	4.0	8.0	12.0	15.0	18.0	21.0	25.0	2.0	7.3	11.5	15.3	18.8	22.1	25.3	1.29	15.27	0.36
8	23,782	6.0	10.0	14.0	17.0	20.0	23.0	27.0	4.0	8.9	13.1	16.9	20.6	24.1	27.5	1.25	16.94	0.33
9	22,608	7.0	12.0	15.0	18.0	22.0	25.0	29.0	5.8	10.3	14.5	18.4	22.1	25.8	29.3	1.21	18.40	0.31
10	21,770	8.0	13.0	16.0	20.0	23.0	26.0	30.0	7.3	11.7	15.7	19.6	23.4	27.0	30.6	1.17	19.62	0.29
11	21,159	10.0	14.0	17.0	21.0	24.0	27.0	31.0	8.6	12.8	16.8	20.6	24.4	28.1	31.7	1.14	20.62	0.28
12	12,941	10.0	15.0	18.0	21.0	25.0	28.0	32.0	9.7	13.7	17.6	21.5	25.2	29.0	32.7	1.09	21.47	0.27
13	15,555	12.0	15.0	19.0	22.0	26.0	29.0	33.0	10.6	14.5	18.4	22.2	26.0	29.8	33.6	1.03	22.20	0.26
14	13,681	12.0	16.0	19.0	23.0	27.0	30.0	34.0	11.3	15.1	18.9	22.8	26.6	30.5	34.4	0.97	22.76	0.25
15	10,135	13.0	17.0	20.0	23.0	27.0	30.0	35.0	11.8	15.5	19.3	23.1	27.0	31.0	35.0	0.91	23.11	0.25
16	7688	13.0	16.0	19.0	23.0	27.0	30.0	34.0	12.1	15.7	19.4	23.3	27.3	31.3	35.5	0.84	23.31	0.25
17	2753	12.0	16.0	19.0	23.0	27.0	30.0	35.0	12.3	15.8	19.5	23.4	27.4	31.6	35.9	0.77	23.40	0.25
18	367	10.0	15.0	18.0	22.0	26.0	30.0	36.0	12.4	15.8	19.5	23.4	27.5	31.7	36.2	0.70	23.40	0.26
<i>10 × 5 m SRT (sec)</i>																		
6	21,133	32.3	28.4	26.1	24.4	22.3	21.0	19.5	31.6	29.0	26.5	24.0	21.4	18.9	16.3	1.0	24.04	0.15
7	27,340	30.0	27.0	25.0	23.0	21.3	20.0	18.6	30.6	28.2	25.7	23.3	20.8	18.3	15.9	1.0	23.32	0.15
8	23,231	29.0	26.0	23.9	22.1	20.5	19.2	18.0	29.7	27.4	25.0	22.6	20.2	17.8	15.4	1.0	22.62	0.15
9	22,143	27.9	25.0	23.1	21.5	20.0	18.7	17.4	28.9	26.6	24.2	22.0	19.6	17.3	15.0	1.0	21.96	0.16
10	25,561	27.0	24.3	22.6	21.0	19.4	18.1	17.0	28.1	25.9	23.6	21.4	19.1	16.8	14.6	1.0	21.37	0.16
11	29,745	26.1	23.8	22.00	20.4	19.0	17.7	16.4	27.5	25.3	23.1	20.9	18.6	16.4	14.2	1.0	20.86	0.16
12	12,695	26.0	23.4	21.8	20.1	18.7	17.3	16.1	27.0	24.8	22.6	20.4	18.3	16.1	14.0	1.0	20.45	0.16
13	15,168	26.3	23.2	21.3	19.8	18.3	16.9	15.7	26.6	24.4	22.3	20.1	18.0	15.9	13.7	1.0	20.13	0.16
14	13,244	26.0	23.0	21.0	19.3	17.8	16.5	15.3	26.3	24.2	22.1	19.9	17.8	15.7	13.6	1.0	19.91	0.16
15	9528	26.5	23.0	21.0	19.2	17.8	16.4	15.0	26.1	24.0	21.9	19.8	17.7	15.6	13.5	1.0	19.77	0.17
16	7146	27.8	23.5	21.1	19.3	17.9	16.4	15.0	26.1	24.0	21.9	19.7	17.7	15.6	13.5	1.0	19.68	0.17
17	2487	29.4	23.9	21.2	19.4	17.8	16.4	5.0	26.0	23.9	21.8	19.6	17.6	15.6	13.5	1.0	19.64	0.17
18	345	31.0	25.8	22.1	19.9	18.2	16.7	15.0	26.0	23.9	21.8	19.6	17.6	15.6	13.5	1.0	19.61	0.17

Note: *P*, percentile; *L*, skew; *M*, median; *S*, coefficient of variation; age: completed age, e.g. 6 years = 6.00–6.99 years.

Table II. Girls physical fitness tests percentiles and LMS values and LMS summary statistics by age in 6–18-year-old Greek children and adolescents

Age	N	Percentiles							LMS method									
		P ₃	P ₁₀	P ₂₅	P ₅₀	P ₇₅	P ₉₀	P ₉₇	P ₃	P ₁₀	P ₂₅	P ₅₀	P ₇₅	P ₉₀	P ₉₇	L	M	S
20 m SRT (completed laps)																		
6	1640	2.0	4.0	8.0	12.0	18.0	26.0	38.0	3.5	5.7	8.8	13.0	18.6	25.9	35.1	0.25	13.02	0.56
7	14,777	5.0	8.0	11.0	15.0	21.0	29.0	39.0	4.1	6.8	10.5	15.4	21.9	30.2	40.6	0.28	15.45	0.55
8	20,084	7.0	9.0	12.0	18.0	25.0	34.0	45.0	4.9	8.0	12.4	18.1	25.6	35.0	46.5	0.30	18.14	0.54
9	19,806	7.0	10.0	14.0	20.0	29.0	40.0	51.0	5.6	9.2	14.3	20.9	29.3	39.8	52.6	0.32	20.87	0.54
10	19,143	8.0	11.0	16.0	23.0	33.0	43.0	55.0	6.1	10.3	16.0	23.4	32.8	44.3	58.2	0.35	23.42	0.54
11	18,290	9.0	12.0	18.0	26.0	36.0	48.0	61.0	6.4	11.1	17.3	25.4	35.6	48.0	62.8	0.37	25.44	0.54
12	10,693	7.0	12.0	18.0	26.0	37.0	49.0	62.0	6.4	11.3	17.9	26.5	37.2	50.1	65.4	0.40	26.49	0.54
13	6767	6.0	12.0	18.0	26.0	37.0	50.0	62.0	6.0	11.0	17.9	26.7	37.6	50.7	66.2	0.42	26.71	0.55
14	5563	5.0	12.0	18.0	26.0	37.0	50.0	63.0	5.5	10.6	17.5	26.4	37.4	50.5	65.9	0.45	26.40	0.56
15	2576	7.0	12.0	17.0	25.0	35.0	47.0	62.0	4.9	9.9	16.9	25.8	36.8	49.9	65.1	0.47	25.82	0.58
16	1462	7.0	11.0	16.0	24.0	33.0	45.0	60.0	4.2	9.2	16.2	25.2	36.2	49.2	64.3	0.49	25.21	0.60
17	490	5.0	10.0	16.0	23.0	36.0	50.0	67.0	3.5	8.5	15.6	24.6	35.7	48.7	63.6	0.52	24.63	0.61
18	240	3.0	8.0	14.0	23.0	28.0	39.5	63.7	2.8	7.8	14.8	23.9	34.9	47.8	62.5	0.54	23.92	0.63
SLJ (cm)																		
6	17,152	59.7	70.0	80.0	93.0	105	116	128	55.3	68.4	81.0	93.2	105	117	128	1.22	93.2	0.19
7	23,303	65.0	78.0	90.0	101	115	126	140	61.5	75.6	89.1	102	115	128	140	1.19	102.3	0.19
8	23,099	73.0	85.0	97.0	110	125	138	150	67.7	82.6	97.1	111	125	139	152	1.16	111.3	0.19
9	22,427	80.0	92.5	105	120	134	148	160	73.6	89.3	105	120	135	149	164	1.13	119.8	0.19
10	21,589	85.0	99.0	111	127	142	156	170	78.5	95.0	111	127	143	158	174	1.10	127.1	0.19
11	20,678	90.0	103	118	134	150	165	180	82.1	99.1	116	133	149	165	182	1.08	132.6	0.19
12	12,171	90.0	105	120	136	153	170	185	83.8	101	119	136	153	170	187	1.07	135.8	0.19
13	7795	90.0	104	120	137	154	170	187	84.0	102	119	137	154	172	189	1.06	137.0	0.19
14	6454	89.0	103	120	135	154	170	188	83.1	101	119	137	155	172	190	1.05	136.9	0.19
15	2058	86.0	100	116	134	151	170	186	81.6	100	118	136	154	172	190	1.05	136.3	0.20
16	7677	86.0	100	115	132	150	170	187	79.9	98.7	117	136	154	172	190	1.06	135.7	0.20
17	2637	83.0	100	116	133	152	170	190	78.1	97.3	116	135	154	172	190	1.07	135.0	0.21
18	240	81.0	99.0	112	130	150	169	197	76.2	95.8	115	134	153	172	190	1.08	134.3	0.21
SR (cm)																		
6	20,463	3.0	7.0	12.0	16.0	20.0	24.0	28.0	3.3	7.4	11.7	16.0	20.5	25.0	29.6	0.93	16.04	0.41
7	26,851	3.0	7.0	12.0	16.0	21.0	25.0	29.0	3.1	7.3	11.6	16.1	20.7	25.4	30.1	0.92	16.10	0.42
8	25,849	3.0	7.0	12.0	16.0	21.0	25.0	30.0	2.9	7.1	11.6	16.2	20.9	25.7	30.7	0.91	16.17	0.43
9	25,623	2.0	6.0	12.0	16.0	21.0	26.0	30.0	2.8	7.0	11.6	16.3	21.1	26.1	31.2	0.90	16.27	0.44
10	24,913	2.0	6.0	11.0	17.0	22.0	26.0	30.0	2.8	7.0	11.7	16.5	21.6	26.7	32.0	0.89	16.54	0.45
11	24,156	2.0	7.0	12.0	17.0	23.0	28.0	32.0	2.8	7.3	12.1	17.1	22.4	27.8	33.3	0.88	17.11	0.45
12	15,228	3.0	7.0	13.0	18.0	24.0	29.0	34.0	3.0	7.6	12.6	17.9	23.4	29.0	34.8	0.87	17.88	0.45
13	14,014	3.0	8.0	14.0	19.0	25.0	30.0	35.0	3.2	8.0	13.2	18.6	24.3	30.2	36.2	0.87	18.63	0.45
14	12,543	3.0	9.0	15.0	20.0	26.0	30.0	36.0	3.4	8.3	13.6	19.3	25.1	31.1	37.3	0.87	19.25	0.45
15	9135	3.0	9.0	15.0	20.0	26.0	31.0	36.0	3.6	8.6	14.0	19.7	25.6	31.6	37.8	0.88	19.68	0.44
16	7539	4.0	9.0	15.0	20.0	26.0	31.0	36.0	3.9	8.9	14.3	19.9	25.8	31.8	38.0	0.88	19.93	0.43
17	2597	4.0	10.0	15.0	20.0	26.0	30.0	35.0	4.0	9.1	14.5	20.1	25.9	31.9	38.0	0.89	20.09	0.43
18	240	3.0	8.0	12.0	18.0	25.0	29.0	31.0	4.2	9.3	14.6	20.2	26.0	31.9	37.9	0.89	20.23	0.42

<i>SUs (no. in 30 sec)</i>																		
6	17,044	2.0	5.0	9.0	12.0	15.0	18.0	22.0	2.0	5.0	9.5	13.2	16.7	19.9	22.9	1.36	13.24	0.41
7	23,158	4.0	8.0	11.0	14.0	18.0	21.0	24.0	2.1	6.5	10.9	14.9	18.5	21.9	25.1	1.33	14.86	0.38
8	23,030	5.0	9.0	13.0	16.0	19.0	22.0	26.0	2.5	7.9	12.4	16.4	20.1	23.6	27.0	1.29	16.35	0.35
9	22,368	6.0	10.0	14.0	17.0	21.0	24.0	28.0	4.4	9.3	13.6	17.6	21.4	25.0	28.4	1.26	17.61	0.33
10	21,511	8.0	12.0	15.0	18.0	22.0	25.0	28.0	5.9	10.5	14.7	18.6	22.3	25.9	29.4	1.22	18.58	0.31
11	20,612	9.0	13.0	16.0	19.0	22.0	25.0	30.0	7.2	11.4	15.4	19.2	22.9	26.5	30.0	1.17	19.21	0.29
12	15,324	10.0	13.0	16.0	19.0	22.0	26.0	30.0	8.0	12.0	15.8	19.5	23.1	26.7	30.2	1.12	19.52	0.28
13	14,181	10.0	13.0	16.0	19.0	22.0	26.0	30.0	8.5	12.3	16.0	19.6	23.1	26.7	30.2	1.07	19.57	0.28
14	12,637	10.0	13.0	16.0	19.0	23.0	26.0	30.0	8.8	12.3	15.9	19.4	23.0	26.5	30.0	1.01	19.42	0.27
15	9264	10.0	13.0	16.0	19.0	22.0	26.0	30.0	8.8	12.2	15.6	19.1	22.6	26.1	29.6	0.96	19.09	0.27
16	7686	9.0	12.0	15.0	18.0	21.0	25.0	29.0	8.6	11.9	15.3	18.7	22.1	25.6	29.2	0.91	18.66	0.28
17	2634	9.0	12.0	15.0	18.0	21.0	25.0	28.0	8.4	11.6	14.8	18.2	21.6	25.1	28.7	0.87	18.18	0.28
18	242	7.0	10.0	14.0	17.0	20.0	24.0	28.0	8.2	11.2	14.4	17.7	21.0	24.5	28.1	0.82	17.65	0.28
<i>10 × 5 m SRT (sec)</i>																		
6	20,030	32.0	29.0	26.9	24.9	23.2	21.8	20.2	32.6	29.9	27.4	25.0	22.7	20.6	18.5	0.42	24.99	0.14
7	26,271	30.3	27.8	25.8	24.0	22.2	20.9	19.2	31.4	28.9	26.4	24.1	21.9	19.8	17.9	0.42	24.10	0.14
8	25,450	29.7	26.7	24.8	23.0	21.4	20.0	18.5	30.3	27.9	25.5	23.3	21.2	19.2	17.3	0.42	23.27	0.14
9	25,394	28.3	25.8	24.0	22.3	20.7	19.4	18.0	29.4	27.0	24.7	22.6	20.5	18.6	16.7	0.42	22.56	0.14
10	24,749	27.8	25.2	23.4	21.8	20.2	18.9	17.5	28.7	26.3	24.1	22.0	20.0	18.1	16.3	0.42	22.01	0.14
11	23,862	27.1	24.6	22.9	21.2	19.7	18.4	17.0	28.2	25.9	23.7	21.6	19.7	17.8	16.1	0.42	21.64	0.14
12	14,920	27.0	24.6	22.9	21.2	19.6	18.3	17.0	27.9	25.7	23.5	21.4	19.5	17.7	15.9	0.42	21.44	0.14
13	13,569	27.8	24.9	23.0	21.3	19.7	18.2	16.9	27.9	25.6	23.4	21.4	19.4	17.6	15.9	0.42	21.39	0.14
14	12,088	28.0	25.0	23.0	21.2	19.6	18.2	17.0	28.0	25.7	23.5	21.4	19.5	17.7	15.9	0.42	21.45	0.14
15	8575	28.3	25.4	23.4	21.5	19.9	18.3	16.8	28.1	25.8	23.6	21.5	19.6	17.7	16.0	0.42	21.56	0.14
16	7004	30.0	26.1	24.0	22.0	20.0	18.4	16.8	28.3	26.0	23.8	21.7	19.7	17.9	16.1	0.42	21.70	0.14
17	2392	30.1	26.1	24.0	22.0	20.0	18.5	16.8	28.5	26.2	23.9	21.8	19.9	18.0	16.2	0.42	21.85	0.14
18	204	32.0	27.7	24.7	22.4	20.2	18.7	16.5	28.7	26.3	24.1	22.0	20.0	18.1	16.3	0.42	22.02	0.14

Note: *P*, percentile; *L*, skew; *M*, median; *S*, coefficient of variation; age: completed age, e.g. 6 years = 6.00–6.99 years.

2011), Portugal (Santos et al., 2014), and Australia (Catley & Tomkinson, 2013) that have examined similar physical fitness tests in children aged 6–18 years. In all the aforementioned studies, boys performed better than girls in cardiorespiratory endurance, speed/agility, muscular strength, and muscular endurance tests, while older ages, in both sexes, have incorporated higher percentile values in comparison with younger ones (Catley & Tomkinson, 2013; Santos et al., 2014; Sauka et al., 2011). Moreover, it seems that physical fitness test performance tends to peak from about the age of 15 years, especially in girls. This finding is in accordance with results from a previous large European epidemiological study (the HELENA study) which found stability in girls' performance in aerobic fitness, speed/agility, and flexibility tests after about the age of 15 years (Ortega et al., 2011).

A finding that deserves attention is the lack of significant differences between the two methods used to create the percentiles in 20 m SRT, SLJ, SR, SUs, and 10 × 5 m SRT, in both sexes. The LMS method uses smoothed curves to estimate the critical centiles and has been extensively been used to provide a way of obtaining normalised growth centiles standards (Cole, 1990). To the best of our knowledge, this is one of the few times that the LMS method has been used to calculate fitness centiles in children and adolescents (Bustamante, Beunen, & Maia, 2012; Catley & Tomkinson, 2013; Eisenmann et al., 2011; Gúliás-González, Sánchez-López, Olivas-Bravo, Solera-Martínez, & Martínez-Vizcaíno, 2014; Santos et al., 2014; Silva, Aires, Mota, Oliveira, & Ribeiro, 2012). None of the above-mentioned studies have made comparisons between two methods. In our study, for at first time we have made comparison between the LMS method and the empirical method which showed that the results can be trusted, at least in large data analysis.

School in Greece, as in most developed countries, is the first institution providing opportunities for physical activity through the pursuit of the physical education curriculum (Cavill, Kahlmeier, & Racioppi, 2006). Given that our study was based on a yearly nationwide school-based survey programme under the auspices of the Ministry of Education, our findings may be of assistance to physical fitness educators and policy-makers when assessing schools' physical education programmes. In fact, it is an opportunity for a significant intervention focusing on elementary and secondary school children physical fitness through the relatively easy access of schools, where changes in physical fitness can be easier implemented and monitored, and affect almost all schoolchildren.

Strengths and limitations

The present study has several strengths. The sample is representative of the national gender and geographical representation, as we studied almost 40% of the total population of 6–18-year-old children in Greece. The study was performed in children aged 6–18 years. This age range is an advantageous period in the life of children and adolescents at which to apply effective prevention strategies to improve physical fitness levels. Finally, the presented data were derived using the same standardized procedures in all schools.

There are also limitations in our study design. Although a common, validated protocol was used to evaluate fitness tests in all schools, a large number of experienced, professional physical educators participated as evaluators in the study. In order to minimize the variability among the different experimenters, all educators were instructed through a detailed and extended manual of operations and followed a standardized procedure of measurements. Even so, some variability in measurement will still exist. Moreover, the 20 m SRT is an objective test of aerobic fitness which indirectly infers peak VO_2 and does not directly measure aerobic fitness. In order to fully understand the development of $\text{VO}_{2\text{max}}$ during adolescence, longitudinal studies are required to take into consideration the tempo and timing of growth and maturation, particularly in girls (Eisenmann, Laurson, & Welk, 2011). Finally, the cross-sectional design of our study cannot provide causal relationships, but only provides hypotheses for further research.

Conclusions

In conclusion, we established sex- and age-specific normative physical fitness values for children aged 6–18 years living in Greece. Boys performed better in all measurements except flexibility than girls of the same age and older children performed better than younger ones. These findings may help policy-makers to design appropriate health-related educational and physical fitness programmes for the young, to facilitate future more detailed epidemiology research on this topic and for comparisons among the same health-related fitness scores of children from other countries similar to Greece.

Acknowledgements

The authors would like to thank the physical education professionals who organized the projects and

performed the evaluations and the students for agreeing to participate. Also, we are grateful to Mrs Sarah Toombs Smith, Ph.D., for assistance with manuscript preparation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Graduate Program, Department of Nutrition and Dietetics of Harokopio University, the Hellenic Ministry of Education and Religions, the Hellenic Ministry of Culture and Athletics, the Secretariat General of Sports, and OPAP SA.

References

- Andersen, L. B., Wedderkopp, N., Hansen, H. S., Cooper, A. R., & Froberg, K. (2003). Biological cardiovascular risk factors cluster in Danish children and adolescents: The European youth heart study. *Preventive Medicine*, 37, 363–367. doi:10.1016/S0091-7435(03)00145-2
- Anderssen, S. A., Cooper, A. R., Riddoch, C., Sardinha, L. B., Harro, M., Brage, S., & Andersen, L. B. (2007). Low cardiorespiratory fitness is a strong predictor for clustering of cardiovascular disease risk factors in children independent of country, age and sex. *European Journal of Cardiovascular Prevention & Rehabilitation*, 14, 526–531. doi:10.1097/HJR.0b013e328011efc1
- Bustamante, A., Beunen, G., & Maia, J. (2012). Evaluation of physical fitness levels in children and adolescents: Establishing percentile charts for the central region of Peru. *Revista Peruana de Medicina Experimental y Salud Publica*, 29, 188–97.
- Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Representative*, 100, 126–131.
- Catley, M. J., & Tomkinson, G. R. (2013). Normative health-related fitness values for children: Analysis of 85347 test results on 9–17-year-old Australians since 1985. *British Journal of Sports Medicine*, 47, 98–108. doi:10.1136/bjsports-2011-090218
- Cavill, N., Kahlmeier, S., & Racioppi, F. (2006). What can the health sector and others do to increase physical activity? In N. Cavill, S. Kahlmeier, & F. Racioppi (Eds.), *Physical activity and health in Europe: Evidence for action* (pp. 15–23). Copenhagen: World Health Organization Regional Office for Europe.
- Christodoulos, A. D., Douda, H. T., & Tokmakidis, S. P. (2012). Cardiorespiratory fitness, metabolic risk, and inflammation in children. *International Journal of Pediatrics*, 12, Article no. 270515. doi:10.1155/2012/270515
- Cole, T. J. (1990) The LMS method for constructing normalized growth standards. *European Journal of Clinical Nutrition*, 44, 45–60. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2354692>
- Cole, T. J., & Green, P. J. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine*, 11, 1305–1319. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1518992>
- Eisenmann, J. C., Laurson, K. R., & Welk, G. J. (2011). Aerobic fitness percentiles for U.S. adolescents. *American Journal of Preventive Medicine*, 41, S106–S110. doi:10.1016/j.amepre.2011.07.005
- Eisenmann, J. C., Wickel, E. E., Welk, G. J., & Blair, S. N. (2005). Relationship between adolescent fitness and fatness and cardiovascular disease risk factors in adulthood: The aerobics center longitudinal study (ACLS). *American Heart Journal*, 149, 46–53. doi:10.1016/j.ahj.2004.07.016
- Ekelund, U., Anderssen, S. A., Froberg, K., Sardinha, L. B., Andersen, L. B., Brage, S., & European Youth Heart Study Group. (2007). Independent associations of physical activity and cardiorespiratory fitness with metabolic risk factors in children: The European youth heart study. *Diabetologia*, 50, 1832–1840. doi:10.1016/S0140-6736(06)69075-2
- Flouris, A. D., Bouziotas, C., Christodoulos, A. D., & Koutedakis, Y. (2008). Longitudinal preventive-screening cutoffs for metabolic syndrome in adolescents. *International Journal of Obesity*, 32, 1506–1512. doi:10.1038/ijo.2008.142
- Gulías-González, R., Sánchez-López, M., Olivas-Bravo, Á., Solera-Martínez, M., & Martínez-Vizcaino, V. (2014). Physical fitness in Spanish schoolchildren aged 6–12 years: Reference values of the battery EUROFIT and associated cardiovascular risk. *Journal of School Health*, 84, 625–635. doi:10.1111/josh.12192
- Hurtig-Wennlof, A., Ruiz, J. R., Harro, M., & Sjöström, M. (2007). Cardiorespiratory fitness relates more strongly than physical activity to cardiovascular disease risk factors in healthy children and adolescents: The European youth heart study. *European Journal of Cardiovascular Prevention & Rehabilitation*, 14, 575–581. doi:10.1097/HJR.0b013e32808c67e3
- Leger, L. A., Lambert, J., Goulet, A., Rowan, C., & Dinelle, Y. (1984). Aerobic capacity of 6 to 17-year-old Quebecois—20 meter shuttle run test with 1 minute stages. *Canadian Journal of Applied Sport Sciences*, 9, 64–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6733834>.
- Leger, L. A., Mercier, D., Gadoury, C., & Lambert, J. (1988). The multistage 20 metre shuttle run test for aerobic fitness. *Journal of Sports Sciences*, 6, 93–101. doi:10.1080/02640418808729800
- Looney, M. A., & Plowman, S. A. (1990). Passing rates of American children and youth on the FITNESSGRAM criterion-referenced physical fitness standards. *Research Quarterly for Exercise and Sport*, 61, 215–223. doi:10.1080/02701367.1990.10608682
- Ortega, F. B., Artero, E. G., Ruiz, J. R., España-Romero, V., Vicente-Rodriguez, G., Molnar, D., ... Gutiérrez, A., HELENA Study Group. (2011). Physical fitness levels among European adolescents: The HELENA study. *British Journal of Sports Medicine*, 45, 20–29. doi:10.1136/bjsm.2009.062679
- Ortega, F. B., Ruiz, J. R., Castillo, M. J., & Sjöström, M. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity*, 32(1), 1–11. doi:10.1038/sj.ijo.0803774
- Pan, H., & Cole, T. J. (2010). LMS Chartmaker, a program to construct growth references using the LMS method. Version 2.43. Retrieved from <http://www.healthforallchildren.co.uk/>
- Pan, H., & Cole, T. J. (2011). LMS growth, a Microsoft Excel add-in to access growth references based on the LMS method. Version 2.71. Retrieved from <http://www.healthforallchildren.co.uk/>
- Santos, R., Mota, J., Santos, D. A., Silva, A. M., Baptista, F., & Sardinha, L. B. (2014). Physical fitness percentiles for Portuguese children and adolescents aged 10–18 years. *Journal of Sports Sciences*, 13, 1–3. doi:10.1080/02640414.2014.906046

- Sauka, M., Priedite, I. S., Artjuhova, L., Larins, V., Selga, G., Dahlström, O., & Timpka, T. (2011). Physical fitness in northern European youth: Reference values from the Latvian physical health in youth study. *Scandinavian Journal of Public Health*, 39, 35–43. doi:10.1177/1403494810380298
- Silva, G., Aires, L., Mota, J., Oliveira, J., & Ribeiro, J. C. (2012). Normative and criterion-related standards for shuttle run performance in youth. *Pediatric Exercise Science*, 24, 157–69.
- Tambalis, K. D., Panagiotakos, D. B., Arnaoutis, G., & Sidossis, L. A. (2013). Endurance, explosive power, and muscle strength in relation to body mass index and physical fitness in Greek children aged 7–10 years. *Pediatric Exercise Science*, 25, 394–406. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23877385>
- Tambalis, K. D., Panagiotakos, D. B., Psarra, G., & Sidossis, L. A. (2011). Inverse but independent trends in obesity and fitness levels among Greek children: A time-series analysis from 1997 to 2007. *Obesity Facts*, 4, 165–174. doi:10.1159/000327994
- Tokmakidis, S. P., Kasambalis, A., & Christodoulos, A. D. (2006). Fitness levels of Greek primary schoolchildren in relationship to overweight and obesity. *European Journal of Pediatrics*, 165, 867–874. doi:10.1007/s00431-006-0176-2
- Wang, Y., Moreno, L. A., Caballero, B., & Cole, T. J. (2006). Limitations of the current world health organization growth references for children and adolescents. *Food and Nutrition Bulletin*, 27, S175–S188. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17361655>

Copyright of European Journal of Sport Science is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.



Short Communication

Gender differences in body fat content are present well before puberty

RW Taylor,¹ E Gold², P Manning² and A Goulding²

¹Department of Human Nutrition and ²Department of Medicine, Otago University, Dunedin, New Zealand.

To determine whether gender differences in body fat could be detected in prepubertal children using dual energy X-ray absorptiometry (DEXA), body composition was measured in 20 healthy boys aged 3–8 y matched for age, height and weight with 20 healthy girls. Although boys and girls did not differ in age, height, weight, body mass index (BMI) or bone mineral content, the boys had a lower percentage of body fat (13.5 ± 5.1 vs $20.4 \pm 6.1\%$, $P < 0.01$), a lower fat mass (3.2 ± 2.0 vs 4.9 ± 3.1 kg, $P < 0.01$), and a higher bone-free lean tissue mass (18.6 ± 4.3 vs 17.0 ± 3.5 kg, $P < 0.01$) than the girls. Girls had approximately 50% more body fat than the boys. This is the first DEXA study to show that boys aged 3–8 y have less body fat than girls of similar age, height and weight. Thus, this technology demonstrates that significant gender differences in body composition are evident, well before the onset of puberty.

Keywords: adiposity; body composition; gender; prepubertal children

Introduction

Adult men are taller, heavier, have a significantly lower percentage of body fat, but a greater lean tissue mass and a higher bone mineral content than women. These gender differences in body fat content are usually said to arise in adolescence. Prepubertally, sexual dimorphism in body composition is considered to be slight¹ and before puberty boys and girls have similar average heights and weights.² However, the precise estimation of body fat content in young children has until recently been difficult. This situation has changed with the advent of dual energy X-ray absorptiometry (DEXA),^{3–5} which provides precise and simple independent measurements of total body fat, lean and bone mineral content in young children.^{6–8} Using this methodology, several recent papers have noted lower body fat in boys than girls aged 9–10 y.^{9,10} The present study was undertaken to test the hypothesis, that DEXA measurements of total body fat content would reveal that, prior to the onset of puberty, even younger boys have a smaller proportion of body fat than girls.

Methods

Participants were healthy volunteers recruited for studies of nutrition and bone health approved by our

hospital Ethics Committee. Their body mass index (BMI) values lay between the 5–95 percentiles of USA reference data.¹¹ All were graded as prepubertal (Tanner Stage I).¹² None were taking medication affecting bone or body composition and no child had previous fractures. Each child was weighed (electronic scales) and measured (stadiometer) in light clothing without shoes. A medical history was taken and a total body DEXA scan was performed (Lunar DPX-L Lunar Corporation, Wisconsin, USA) to measure total fat mass, bone-free lean tissue mass and bone mineral content. Our DEXA precision (coefficient of variation) *in vivo* is: fat mass 2.52%; bone-free lean tissue mass 1.11% and bone mineral content 1.52%. The total percentage fat in the body determined from DEXA measurements was calculated as fat mass (kg) divided by DEXA mass [that is bone free lean tissue mass (kg) plus bone mineral content (kg) plus fat mass (kg)] $\times 100/1$.

The DEXA scan data of a group of healthy boys aged 3–8 y, who were enrolled consecutively in a study of bone development, were matched with scan data from a group of healthy girls collected simultaneously for our DEXA bone density data base. The scan results of each boy were compared with the scan results from the girl who most closely matched them in age, height and body weight. Gender matches obtained in this way were close (Table 1). Results are expressed as means $\pm$ s.d. Ninety-five percent confidence intervals for the differences between each pair were calculated (Mean $\pm$ s.e.m.). Gender differences were tested (Student's paired *t*-test).

Correspondence: Dr Ailsa Goulding, Department of Medicine, Otago University, PO Box 913, Dunedin, New Zealand.
Received 10 June 97; accepted 6 August 97

Results

Although the 20 pairs of prepubertal girls and boys in this study were similar in age, height, weight, BMI and bone mineral content (BMC), boys had significantly less fat, a lower % body fat and a higher bone-free lean tissue mass than the girls (Table 1 and Figure 1). Standard deviation scores¹³ for height and weight respectively, were: 1.06 ± 0.65 and 0.85 ± 0.92 in girls and 0.77 ± 0.84 and 0.88 ± 1.05 in boys. The average fat mass of the girls (4.9 kg), expressed as a percentage of the average fat mass of the boys, was 152% (3.2 kg, $P < 0.05$) while the bone-free lean tissue mass was 9% lower ($P < 0.05$) and the BMC content 2% lower in girls than in boys (not statistically significant $P > 0.05$).

Discussion

This appears to be the first gender comparison of body fat content in tightly age-matched children aged 3–8 y measured by DEXA. Our results provide strong support for the view that boys have a lower fat content than girls, even before puberty, despite similarities in overall body weight, height and BMI. Interestingly, a recent paper⁷ contained some information concerning percentage body fat estimated by DEXA in children aged 4–8 y; values of $19.0 \pm 7.2\%$ body fat in 22 girls and $14.1 \pm 4.1\%$ body fat in 23 boys being reported. However these children were not pair-matched for age, height and weight, and the authors⁷ did not test the gender differences, nor did they report either fat mass or bone-free tissue mass. We have calculated that their young boys had significantly less fat than

the girls ($P < 0.01$), as in our study. Thus their gender difference is of the same order as that we observed, providing independent support for our present results.

Population bias could explain our findings and be a study limitation. However, the weights and heights of our subjects were within normal limits for a representative study population of 1073 local children measured at 3, 5 and 7 y (PA Silva, personal communication). Moreover, international standard deviation scores confirm that our study population is representative.¹³ We acknowledge that the differences we observed, may reflect differences in physical activity, nutrition or metabolism. There is evidence that boys are more active than girls, both prepubertally^{14,15} and later.¹⁶ Gender differences in energy intake have also been noted in young children.¹⁷ Alternatively, gender differences in leptin may be important in favouring early fat accumulation in girls.¹⁸

Others have observed similar total BMC values in boys and girls aged 8–9 y.^{9,10} In our study of younger

Table 1 Body composition characteristics of children ($n = 20$ matched pairs).

Gender		Age (y)	Height (m)	Weight (kg)	BMI (kg/m ²)	Total fat (kg)	Body fat (%)	Lean tissue mass (kg)	Total body BMC (kg)
Girls	Mean $\pm$ s.d.	6.4 ± 1.7	1.19 ± 0.12	23.5 ± 6.1	16.3 ± 1.7	$4.9 \pm 3.1^*$	$20.4 \pm 6.1^*$	$17.0 \pm 3.5^*$	0.78 ± 0.23
Boys	Mean $\pm$ s.d.	6.4 ± 1.6	1.19 ± 0.12	23.6 ± 6.2	16.4 ± 1.4	3.2 ± 2.0	13.5 ± 5.1	18.6 ± 4.3	0.80 ± 0.25
Girls	Range	3.9–8.9	1.01–1.37	15.6–39.0	13.9–20.8	2.5–14.8	12.1–37.2	11.5–23.7	0.49–1.23
Boys	Range	4.0–8.9	0.99–1.42	15.0–39.8	14.5–19.7	1.5–8.2	7.5–28.6	12.2–29.0	0.44–1.34
	Mean Difference ^a	0.04	0.0	–0.02	–0.1	1.7	6.9	–1.5	–0.02
	95% CI	–0.08–0.16	–0.01–0.02	–0.6–0.5	–0.6–0.4	0.7–2.6	4.2–9.5	–2.3–0.8	–0.05–0.02

Analysed by paired t -test. * $P < 0.01$. ^aGirls relative to boys.

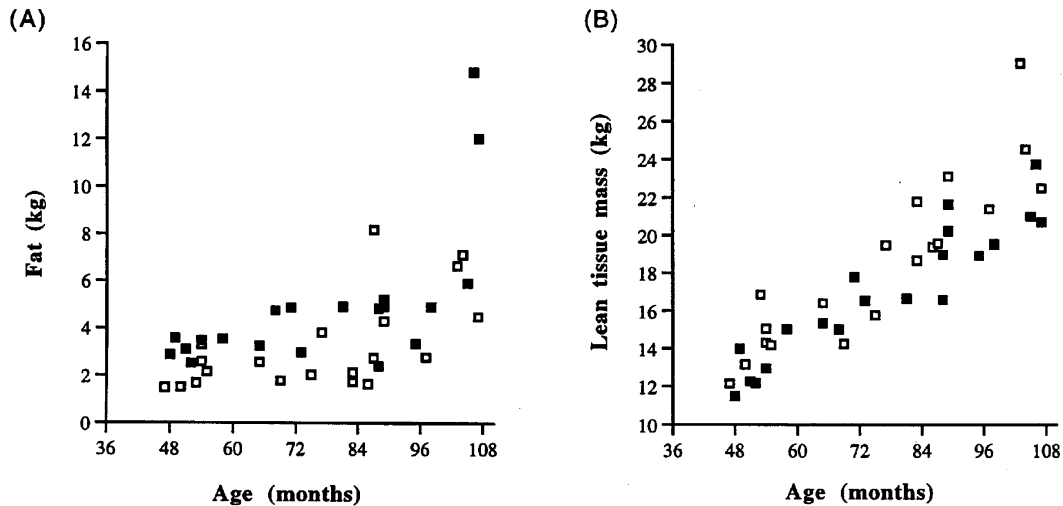


Figure 1 A, Total body fat mass (kg) and B, total body bone-free lean tissue mass (kg), in boys (open squares) and girls (closed squares) in relation to age.

children, BMC was similar in both sexes, providing support for the view that BMC is influenced by total body mass rather than lean tissue or fat tissue.¹⁹

The rising prevalence of obesity is currently a major concern in the western world²⁰ and the ability to quantify fat mass in young children may enable better definition of future health hazards associated with obesity. The common view that prepubertal children have similar body composition, may have arisen due to the lack of suitable methods for the accurate determination of body fat in this age group. With DEXA, we have now shown that significant gender differences in body fat are present in prepubertal children. Our findings need to be confirmed in a larger sample, drawn randomly from the general population.

Acknowledgements

We thank Mrs S Williams, biostatistician in the Department of Preventive and Social Medicine for statistical advice. We gratefully acknowledge funding support from The Otago Medical Research Foundation, The Health Research Council of New Zealand and the NZ Lottery Grant Board.

References

- Forbes G. Human body composition. Growth, aging, nutrition and activity. Springer-Verlag: New York, 1987.
- Kelly J, Stanton W, McGee R, Silva P. Tracking relative weight in subjects studied longitudinally from ages 3 to 13 years. *J Paediatr Child Health* 1992; **28**: 158–161.
- Mazess RB, Barden HS, Bisek JP, Hanson J. Dual-energy x-ray absorptiometry for total body and regional bone-mineral and soft-tissue composition. *Am J Clin Nutr* 1990; **51**: 1106–1112.
- Kohrt WM. Body composition by DXA: tried and true? *Med Sci Sports Med* 1995; **27**: 1349–1353.
- Ellis K. Measuring body fatness in children and young adults: comparison of bioelectrical impedance analysis, total body electrical conductivity, and dual energy x-ray absorptiometry. *Int J Obes* 1996; **20**: 866–873.
- Ogle GD, Allen JR, Humphries IRJ, Lu PW, Briody JN, Morley K, Howman-Giles R, Cowell CT. Body-composition assessment by dual-energy x-ray absorptiometry in subjects aged 4–26 y. *Am J Clin Nutr* 1995; **61**: 746–753.
- Lazarus R, Baur L, Webb K, Blyth F. Body mass index in screening for adiposity in children and adolescents: systematic evaluation using receiver operating characteristic curves. *Am J Clin Nutr* 1996; **63**: 500–506.
- Taylor RW, Cannan R, Gold E, Lewis-Barned N, Goulding A. Regional body fat distribution in New Zealand girls aged 4–16 years: a cross-sectional study by dual energy x-ray absorptiometry. *Int J Obes* 1996; **20**: 763–767.
- Faulkner RA, Bailey DA, Drinkwater DT, Wilkinson AA, Houston CS, McKay HA. Regional and total body bone mineral content, bone mineral density and total body tissue composition in children 8–16 years of age. *Calcif Tiss Int* 1993; **53**: 7–12.
- Nelson D, Simpson P, Johnson C, Barondess D, Kleerekoper M. The accumulation of whole body skeletal mass in third- and fourth-grade children: effects of age, gender, ethnicity and body composition. *Bone* 1997; **20**: 73–79.
- Hammer LD, Kraemer HC, Wilson DM, Ritter PL, Dornbusch SM. Standardized percentile curves of body-mass index for children and adolescents. *Am J Dis Child* 1991; **145**: 259–263.
- Duke PM, Litt IF, Gross RT. Adolescents' self-assessment of sexual maturation. *Pediatrics* 1980; **66**: 918–920.
- Tanner JM, Whitehouse RH. Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. *Arch Dis Child* 1976; **51**: 170–179.
- Gilliam T, Freedson P, Grenen D, Shahraray B. Physical activity patterns determined by heart rate monitoring in 6–7 year-old children. *Med Sci Sports Exerc* 1981; **13**: 65–67.
- Gutin B, Basch C, Shea S, Contento I, DeLozier M, Rips J, Irigoyen M, Zybert P. Blood pressure, fitness, and fatness in 5- and 6-year-old children. *JAMA* 1990; **264**: 1123–1127.
- Armstrong N, Balding J, Gentle P, Kirby B. Patterns of physical activity among 11 to 16 year old British children. *BMJ* 1990; **301**: 203–205.
- Briefel R, McDowell M, Alaimo K, Caughman C, Bischof A, Carroll M, Johnson CL. Total energy of the US population: the third National Health and Nutrition Examination Survey, 1988–1991. *Am J Clin Nutr* 1995; **62** (Suppl): 1072S–1080S.
- Hassink SG, Sheslow DV, de Lancey E, Opentanova I, Considine RV, Caro JF. Serum leptin in children with obesity: relationship to gender and development. 1997; **98**: 201–203.
- Moro M, Van der Meulen M, Kiratli B, Marcus R, Bachrach L, Carter D. Body mass is the primary determinant of mid-femoral bone acquisition during adolescent growth. *Bone* 1996; **19**: 519–526.
- Prentice A, Jebb S. Obesity in Britain: gluttony or sloth? *BMJ* 1995; **311**: 437–439.

Sex Differences in Regional Body Fat Distribution From Pre- to Postpuberty

Rachael W. Taylor¹, Andrea M. Grant¹, Sheila M. Williams² and Ailsa Goulding¹

Few large studies have evaluated the emergence of sexual dimorphism in fat distribution with appropriate adjustment for total body composition. The objective of this study was to determine the timing and magnitude of sex differences in regional adiposity from early childhood to young adulthood. Regional fat distribution was measured using dual-energy X-ray absorptiometry (trunk and extremity fat using automatic default regions and waist and hip fat using manual analysis) in 1,009 predominantly white participants aged 5–29 years. Subjects were divided into pre (Tanner stage 1), early (Tanner stages 2–3), late (Tanner stages 4–5), and post (males ≥ 20 years and females ≥ 18 years) pubertal groups. Sexual dimorphism in trunk fat (adjusted for extremity fat) was not apparent until late puberty, when females exhibited 17% less ($P < 0.001$) trunk fat than males. By contrast, sex differences in waist fat (adjusted for hip fat) were apparent at each stage of puberty, the effect being magnified with age, with prepubertal girls having 5% less ($P = 0.027$) and adult women having 48% less ($P < 0.0001$) waist fat than males. Girls had considerably more peripheral fat whether measured as extremity or hip fat at each stage. Sex differences in regional adiposity were significantly greater in young adults than in late adolescence. Exclusion of overweight participants did not materially affect the estimates. Sexual dimorphism in fat patterning is apparent even prepubertally with girls having less waist and more hip fat than boys. The magnitude of the sex difference is amplified with maturation, and particularly from late puberty to early adulthood.

Obesity (2010) **18**, 1410–1416. doi:10.1038/oby.2009.399

INTRODUCTION

Marked sexual dimorphism in total and regional body composition is apparent in adults, with men having greater lean tissue mass and a more central fat pattern compared with the more peripheral fat distribution typically observed in adult women (1). It is now well accepted that significant gender differences in total body composition are apparent even in prepubertal children (2–4). However, the emergence of sex differences in fat patterning appears to be more controversial. Although anthropometric techniques suggested that sex-specific fat patterning emerges during puberty (5,6), recent work utilizing direct measures of body fat distribution including dual-energy X-ray absorptiometry (DXA) and magnetic resonance imaging has suggested that variation in fat patterning may also exist prepubertally (3,4,7–10). However, the literature is not consistent, with several groups reporting no gender differences in regional adiposity in young children (11–13). Much of this discrepancy can probably be attributed to variations in the regional indexes of choice, the particular ages being studied, differences in relative fatness of the groups, and use of different adjustments for covariates (14).

Few previous groups have studied large samples of participants over the full span of growth from prepuberty to early

adulthood to examine the timing of the appearance of sexual dimorphism in fat distribution (8,11). Elucidating such information is important because of the role of body composition during growth in predicting adult disease risk (15). The aim of this study was to examine when differences in fat distribution measured using anthropometry and DXA become apparent in a large cross-sectional cohort of children and young adults aged 5–29 years.

METHODS AND PROCEDURES

Subjects

Cross-sectional data on total and regional body composition were available for 1,009 predominantly white children and young adults aged 5–29 years who had participated in various studies investigating body composition and health in our laboratory from 1996 to 2007 (16–22). All studies were approved by the Lower South Ethics Committee and written informed consent was obtained from each participant or parent/guardian. A brief medical history was obtained by questionnaire and no participant was taking medication that would affect his or her body composition and no subjects had a history of constitutional delay in growth or maturation. Tanner stage of pubertal development was assessed in all subjects < 19 years of age by the parent/guardian of the younger children or by the adolescent themselves using standard descriptions and pictures (23). Participants were divided into pre

¹Department of Medical and Surgical Sciences, University of Otago, Dunedin, New Zealand; ²Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand. Correspondence: Rachael W. Taylor (rachael.taylor@otago.ac.nz)

Received 19 July 2009; accepted 3 October 2009; published online 5 November 2009. doi:10.1038/oby.2009.399

(Tanner stage 1), early (Tanner stages 2 and 3), late (Tanner stages 4 and 5) (8), and post (males aged ≥ 20 years and females ≥ 18 years) pubertal groups for analysis.

Anthropometric measures

Duplicate measures of height (wall-mounted stadiometer) and weight (electronic scales) were obtained with the participants wearing light clothing and no shoes. Duplicate measures of waist (minimum circumference between the rib cage and the iliac crest using a non-stretchable tape measure) and hip (maximum protuberance of the buttocks) circumferences were obtained in 908 (90%) and 846 (84%) of participants, respectively. Measurements were obtained by several trained examiners following the same protocols. In our laboratory, coefficients of variation (CVs%) for height, weight, and waist circumference in young children are $<2\%$ (17). BMI was calculated as weight in kilograms/height in meters squared. Overweight and obesity were defined using international reference norms (24) for those aged 5–18 years and BMI 25–29.9 and BMI ≥ 30 for overweight and obesity, respectively, for those ≥ 19 years of age.

DXA scanning

All DXA measurements were performed and analyzed by one experienced operator with a Lunar DPX-L scanner (software package 4.7; Lunar, Madison, WI) using standard procedures. The scanner determines total fat mass (kg) and the fat content (kg) of specific anatomical regions including trunk and extremity fat (automatic default regions) and central and peripheral fat (manual regions of interest) as shown in [Figure 1](#). The trunk region consists of the area bordered by a horizontal line below the chin, vertical borders lateral to the ribs, and oblique lines passing through the femoral necks. The arm region consists of all tissues outside these lateral borders and the leg region of all tissue below the oblique lines. Extremity fat was defined as the sum of arm and leg fat (7). Central and peripheral regions of interest were also determined using manual analysis. The central region of interest (waist fat) was a box 9.6 cm high with the lower border positioned superior to the iliac crest. The peripheral region (hip fat) consists of the same sized box, positioned so that the center of the box was at the level of the greater trochanters (25). In our laboratory, the CVs for repeated *in vivo* scans on 10 adults were 2.6% for fat mass, 2.5% for fat percentage, and $<3.5\%$ for all regional measurements.

Statistics

All data were analyzed using STATA (Stata Statistical Software, Release 8.0, 2003; StataCorp, College Station, TX). Analysis of covariance, which adjusted for age and included a sex $\times$ group interaction to test for a different pattern of results across the four different stages of puberty for males and females, was used to analyze the data. Logistic regression was used to analyze the data for overweight status. Regression models were used to examine the sex differences, presented as ratios of female relative to male values, for trunk fat adjusted for extremity fat, and for extremity fat adjusted for trunk fat, for the four pubertal stages. Sex differences for waist fat and hip fat, and waist girth and hip girth were also estimated. All variables not normally distributed (weight, BMI, fat masses) were log transformed before analysis.

RESULTS

[Table 1](#) shows means (s.d.) in total body composition for each pubertal group by sex. Sex differences were observed in all indexes with the exception of the prevalence of overweight/obesity. Furthermore, the interaction between sex and stage indicated that the pattern of differences between males and females was generally higher in each successive stage of puberty. In general, higher values are apparent across puberty for most indexes and for both sexes, with the exception of body fat in

males. Although absolute fat mass is approximately twofold higher in adolescent and older males compared with prepubertal males, relative fatness (fat mass index, % fat) is highest during early puberty. By contrast, females display steady gains in absolute and relative fatness with maturation. Interestingly, the significantly higher body weight of young adults compared with adolescents in late puberty, was not matched by higher stature, and was explained by a higher deposition of lean than fat mass.

Significant differences in regional fat distribution are also apparent with both sex and pubertal development ([Table 2](#)). The greater total body fat of females is also reflected by greater deposition of absolute regional fat at each site with most stages of maturation. In males, higher values are observed for trunk fat mass with age whereas extremity fat mass remains similar from early to postpuberty. Significant interactions between pubertal stage and sex were also observed for most measures showing that the pattern of differences between males and females depended on the level of maturation. By contrast, the nonsignificant interaction effect for waist circumference shows that the sex difference in waist circumference did not differ according to stage of puberty.

[Table 3](#) demonstrates regional fat distribution at each stage of pubertal maturation as ratios (female values relative to males). Interestingly, sexual dimorphism in trunk fat deposition was not apparent until late puberty, and became more marked in young adults, where women had 34% less trunk fat than men, once adjusted for extremity fat. By contrast, differences in waist fat were apparent at each stage of puberty, although this effect magnified with increasing age, ranging from a difference of 5% ($P = 0.027$) in prepubertal children to 48% ($P < 0.001$) in young adults. Considerable differences in peripheral fat distribution were also apparent even in the youngest children. Thus prepubertal girls had 2% ($P = 0.283$) more extremity fat and

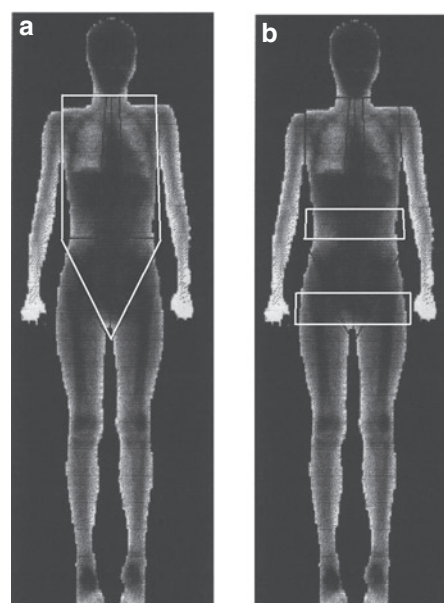


Figure 1 Automatic (a: trunk, arm, and leg) and manual (b: waist and hip) regions of interest.

Table 1 Total body composition by sex and puberty

	Pubertal stage				P		
	Pre	Early	Late	Post	Sex	Stage	Sex × stage
N							
Males	205	92	119	102			
Females	171	97	119	104			
Age (years)							
Males	7.9 (2.3)	13.6 (1.5)	17.0 (1.8)	24.0 (2.6)	<0.001	<0.001	0.002
Females	7.3 (1.7)	12.4 (1.8)	15.8 (1.5)	21.9 (3.2)			
Height (cm)							
Males	127.9 (13.5)	159.6 (10.2)	176.8 (7.1)	179.6 (7.0)	<0.001	<0.001	<0.001
Females	122.9 (11.1)	152.1 (10.0)	164.3 (5.8)	166.4 (5.7)			
Weight (kg)							
Males	29.1 (9.1)	51.8 (11.8)	69.8 (11.4)	79.4 (10.4)	<0.001	<0.001	<0.001
Females	25.9 (6.4)	48.5 (12.6)	63.0 (11.2)	67.0 (14.3)			
BMI (kg/m ²)							
Males	17.4 (2.6)	20.2 (3.4)	22.3 (2.9)	24.6 (2.5)	0.033	<0.001	0.018
Females	17.0 (2.2)	20.8 (4.0)	23.3 (4.0)	24.3 (5.2)			
Overweight/obese (%) ^a							
Males	17/5	15/5	10/3	31/4	0.987	<0.001	0.01
Females	17/5	25/10	24/8	16/13			
Lean mass (kg)							
Males	21.8 (5.5)	37.5 (7.0)	54.8 (7.4)	61.9 (7.0)	<0.001	<0.001	<0.001
Females	18.8 (3.7)	31.5 (5.6)	38.7 (3.6)	41.4 (4.6)			
Fat-free mass index ^b							
Males	13.1 (0.9)	14.6 (1.3)	17.5 (1.5)	19.2 (1.7)	<0.001	<0.001	<0.001
Females	12.3 (0.7)	13.5 (1.1)	14.3 (1.1)	15.0 (1.4)			
Fat mass (kg)							
Males	5.5 (4.6)	11.8 (8.2)	11.7 (7.2)	13.7 (6.3)	<0.001	<0.001	<0.001
Females	5.4 (3.4)	14.6 (8.7)	21.3 (8.9)	22.2 (11.5)			
Fat mass index ^c							
Males	3.2 (2.2)	4.6 (3.2)	3.7 (2.3)	4.2 (1.8)	<0.001	<0.001	<0.001
Females	3.5 (1.8)	6.2 (3.4)	7.9 (3.3)	8.1 (4.3)			
Fat (%)							
Males	17.7 (7.9)	21.5 (10.3)	16.2 (7.0)	16.9 (6.1)	<0.001	<0.001	<0.001
Females	20.4 (6.9)	28.6 (10.4)	33.0 (8.0)	31.9 (9.5)			

Analysis based on analysis of covariance adjusted for age and included terms for sex and pubertal stage and an interaction between sex and pubertal stage. Ordinal logistical regression adjusting for age was used to analyze the variable for overweight and obesity.

^aData are presented as mean (s.d.) except for the prevalence of overweight/obese (24) which is presented as % of group who are overweight but not obese/% of group who are obese. ^bFat-free mass index is lean mass (kg) divided by height squared. ^cFat mass index is fat mass (kg) divided by height squared.

6% ($P < 0.001$) more hip fat than prepubertal boys and these differences became more marked at each successive stage of puberty. Although substantial sex differences in regional adiposity were apparent by late puberty, **Table 3** demonstrates that this dimorphism becomes considerably more marked in young adults. For example, adjusted trunk fat was 17% ($P < 0.001$) and 34% ($P < 0.001$) lower in late and postpubertal females, respectively, relative to males, with corresponding figures of 35% ($P < 0.001$) and 48% ($P < 0.001$) in these

groups for adjusted waist fat. Similar patterns were observed for peripheral fat distribution. Waist circumference was significantly lower and hip circumference significantly higher in girls from early puberty onward.

Exclusion of overweight participants did not materially affect these estimates. The adjusted ratios at each pubertal stage were similar in the total sample ($N = 1,009$) and when analyses were restricted to the normal-weight participants ($N = 846$) for trunk or extremity fat (**Figure 2**) and waist or hip fat (data not shown).

Table 2 Regional body composition by sex and puberty

	Pubertal stage				P		
	Pre	Early	Late	Post	Sex	Stage	Sex × stage
<i>N</i>							
Males	205	92	119	102			
Females	171	97	119	104			
Trunk fat (kg)							
Males	2.0 (2.2)	4.8 (3.9)	5.2 (3.6)	6.8 (3.3)	<0.001	<0.001	<0.001
Females	1.9 (1.6)	6.1 (4.2)	9.2 (4.4)	9.7 (5.8)			
Extremity fat (kg)							
Males	2.9 (2.2)	6.2 (4.1)	5.9 (3.5)	6.3 (3.0)	<0.001	<0.001	<0.001
Females	2.8 (1.7)	7.6 (4.4)	11.2 (4.5)	11.4 (5.4)			
Trunk fat (% total fat)							
Males	31.6 (5.6)	38.6 (5.6)	43.3 (5.3)	49.1 (4.2)	0.033	<0.001	<0.001
Females	34.1 (5.3)	39.9 (5.8)	42.4 (4.1)	42.3 (5.4)			
Trunk-to-leg fat ratio							
Males	0.73 (0.19)	0.86 (0.21)	1.04 (0.25)	1.36 (0.28)	<0.001	<0.001	<0.001
Females	0.77 (0.19)	0.91 (0.23)	0.97 (0.18)	0.99 (0.24)			
Waist fat (kg)							
Males	0.5 (0.5)	0.9 (0.8)	0.8 (0.7)	1.0 (0.6)	<0.001	<0.001	<0.001
Females	0.4 (0.4)	1.1 (0.8)	1.5 (0.8)	1.6 (1.1)			
Hip fat (kg)							
Males	0.8 (0.5)	1.4 (0.8)	1.3 (0.6)	1.4 (0.6)	<0.001	<0.001	<0.001
Females	0.9 (0.4)	1.8 (0.8)	2.5 (0.8)	2.6 (1.1)			
Waist girth (cm)							
Males	59.6 (8.2)	72.4 (9.8)	78.6 (8.1)	83.9 (7.7)	<0.001	<0.001	0.766
Females	56.6 (5.5)	69.6 (9.3)	74.0 (8.8)	79.0 (13.6)			
Hip girth (cm)							
Males	71.4 (8.9)	88.0 (8.6)	96.7 (6.6)	98.9 (6.7)	0.013	<0.001	<0.001
Females	66.9 (6.6)	87.6 (10.2)	98.3 (8.1)	101.1 (11.7)			
Waist-to-hip girth ratio							
Males	0.85 (0.05)	0.82 (0.06)	0.81 (0.05)	0.85 (0.07)	<0.001	<0.001	<0.001
Females	0.85 (0.05)	0.80 (0.06)	0.75 (0.05)	0.78 (0.08)			

Data are presented as mean (s.d.). Values for *n* for waist circumference are 202, 90, 110, and 59 and 166, 96, 113, and 72 for pre, early, late, and post stages of puberty in boys and girls, respectively. Values for *n* for hip circumference and waist-to-hip ratio are as for waist except in the prepubertal group, where *N* = 168 in boys and *N* = 138 in girls. Analysis based on analysis of covariance adjusted for age and included terms for sex and pubertal stage and an interaction between sex and pubertal stage.

DISCUSSION

Our data demonstrate that sexual dimorphism in regional fat distribution becomes apparent at different stages of puberty, depending on the index chosen. Significant sex differences in trunk fat were not evident until late puberty, whereas even in prepubertal children, girls had significantly less waist fat than boys. Differences in peripheral fat were also evident at younger ages, whether measured using extremity fat (sum of leg and arm fat) or hip fat (fat directly over the greater trochanters). Sex differences in waist and hip circumferences were also observed from early puberty. The magnitude of the sex differences in regional adiposity increased significantly with maturation, so

that striking differences were evident between adolescents in late puberty and young adults in their 20s.

The precise timing at which sex differences in regional adiposity is said to become apparent during growth is conflicting in the literature, due in part to the variety of methods used to assess centrality and to choice of statistical adjustments for other body components. In general, studies utilizing skinfold measurements have demonstrated that although absolute thicknesses are often higher in females at least from 8 years of age onward (6,26,27), demarcation in relative skinfold thickness occurs during puberty, primarily due to increased deposition of peripheral fat in females (5,6). Sexual dimorphism

Table 3 Sex differences (as ratios) in various measures of regional fat distribution across puberty (females relative to males)

	Adjusted ratios (95% CI)			
	Pre	Early	Late	Post
Trunk fat (adjusted for extremity fat)	1.01 (0.97, 1.06)	0.99 (0.94, 1.06)	0.83 (0.80, 0.88)	0.66 (0.62, 0.71)
Extremity fat (adjusted for trunk fat)	1.02 (0.98, 1.05)	1.06 (1.01, 1.11)	1.31 (1.25, 1.37)	1.53 (1.46, 1.61)
Waist fat (adjusted for hip fat)	0.95 (0.91, 0.99)	0.85 (0.80, 0.91)	0.65 (0.61, 0.69)	0.52 (0.49, 0.56)
Hip fat (adjusted for waist fat)	1.06 (1.03, 1.09)	1.16 (1.11, 1.21)	1.47 (1.41, 1.53)	1.66 (1.59, 1.73)
Waist girth (adjusted for hip girth)	0.99 (0.98, 1.01)	0.95 (0.94, 0.97)	0.90 (0.89, 0.92)	0.89 (0.87, 0.91)
Hip girth (adjusted for waist girth)	1.00 (0.99, 1.01)	1.04 (1.03, 1.06)	1.09 (1.08, 1.11)	1.11 (1.09, 1.12)

Data are presented as ratios (95% confidence interval (CI)) in females relative to males adjusted for height, age, sex, pubertal stage, interactions between sex and pubertal stage, and specific regional fat as shown.

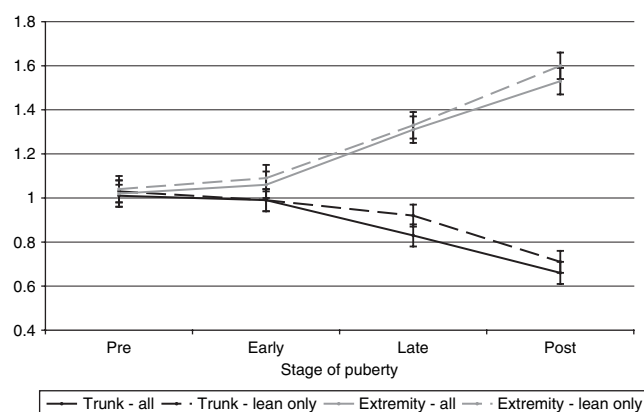


Figure 2 Ratios (females relative to males) for adjusted trunk fat (dark lines) and adjusted extremity fat (light lines) for the total group (solid lines) and lean subjects only (dashed lines).

in circumference measures is also apparent from a young age with elevated waist circumference and/or waist-to-hip ratio or lower hip circumferences being reported in boys as young as 5–7 years (4,26,28,29). Our data support this earlier work, demonstrating that females have lower waist and greater hip circumferences from early puberty. However, there are reservations regarding the use of anthropometric measures to provide indexes of regional adiposity, because of variation in body proportions with growth (15). The advent of DXA, magnetic resonance imaging, and computed tomography provide more accurate and sensitive evaluation techniques to assess changes in body composition, even in young children (15).

Central fat distribution assessed by DXA has typically been reported using trunk fat as a percentage of total fat (% trunk fat) or the trunk-to-leg fat ratio. Although two groups found that differences in trunk-to-leg fat ratio were not apparent until after puberty (11,30), others (4) reported higher trunk-to-leg fat ratio values in 7–8-year-old girls compared with boys (4). Conversely, Wang *et al.* (10) in a large sample of Chinese children, reported that 6–11-year-old boys have a higher % trunk fat than girls. Regardless of these discrepant findings, alternative methods of analyzing DXA data are warranted given the problems of using ratios in statistical analyses (31) and the difficulties in interpreting measures of adiposity which include the measure of interest (e.g., trunk fat) in both the numerator

and the denominator (% trunk fat) (32). Comparing regional fat masses without making adjustments for body size is clearly inappropriate (15); girls will have more body fat in every region as they develop, given their considerably higher total body fat levels and the strong correlations which are present between regional fat and total fat. Thus as recommended recently (15), we adjusted specific central regions for their corresponding peripheral region, as well as adjusting for height, but not for weight. When we included weight as a covariate as other groups have done (7,8), the variance inflation factors indicated that multicollinearity between variables was an issue.

Our results support and extend earlier work (7,8) demonstrating that sexual dimorphism is apparent not only in peripheral fat but also in central fat during puberty. Although expected differences in trunk fat were not apparent until late puberty, we observed that even young prepubertal girls stored less fat in the waist region and more fat peripherally (over the hips) than boys. Use of the DXA manual “waist” region of interest, which focuses narrowly on fat located in the mid-abdominal regional of the body, may provide a more appropriate measure of “central” fat than is provided by the larger default trunk fat region (33), given that DXA cannot differentiate between visceral and subcutaneous fat. However, others have demonstrated that trunk fat measured by DXA is strongly correlated with visceral fat in young children (34). Furthermore, central fat (DXA) is similarly associated with adverse plasma lipid profiles in children as is visceral adipose tissue (35,36). DXA is also a more available and less invasive method of assessing body composition in children, compared with computed tomography or magnetic resonance imaging, which are more costly, take longer to complete and have a larger radiation dose (computed tomography only). Visceral fat accumulation is thought to be similar in prepubertal boys and girls (3,12), particularly after appropriate adjustment for subcutaneous fat (13), although not all studies are consistent (9,37). On the other hand, in prepubertal girls of this age, more of this central fat may be subcutaneous (3).

Interestingly, we observed greater sexual dimorphism in regional adiposity in young adults aged 18–29 years when compared with adolescents in late puberty. Much of this response was probably driven by the differences we observed between adolescent and young adult males; the latter had 2 kg more

total fat and significantly more trunk fat, but similar levels of extremity fat relative to adolescent males. Similar results have been reported recently in a longitudinal study of 17-year-old males followed for 8 years; here both active and inactive men became fatter and stored more of that fat abdominally by study end (38).

Our study is cross-sectional and therefore cannot determine actual changes in body composition within individuals. Pubertal status was self-reported rather than examined, and others (39) have reported only moderate concordance between self- and physician-based assessment. However, Duke *et al.* (23) show that 86–91% of adolescents correctly estimate their developmental age. We also do not have measures of blood hormone status, such as insulin-like growth factor I, sex steroids, and leptin which might explain variation in body composition over time and between the sexes (40,41). Others (4) have shown that variation in insulin-like growth factor I and sex steroids explain <18% of the variance in body composition in prepubertal children. However, this may underestimate the contribution of sex hormones, given that blood sampling in Garnett *et al.* (4) occurred in the early morning, and sex steroids peak during the night. Although recent research has provided greater understanding of the role of sex hormones on fat partitioning, considerable work is also required to determine the role of chromosomal sex (XX, XY) and prenatal hormones in determining fat patterning later in life (42).

Strengths of our study include measurements on a large database of children and young adults varying markedly in shape and size. Measurements were made on the same DXA scanner, and analysis of all DXA data was completed by the same person. In addition, we excluded participants with any known endocrine disorders or with a history of constitutional delay in growth or maturation from our study.

In conclusion, our study shows significant sexual dimorphism in fat patterning is apparent even prepubertally, with girls having less waist and more hip fat than boys. The magnitude of the sex difference is amplified with maturation, with young adults displaying considerably higher central fat deposition compared with those in late adolescence. Progression to young adulthood may be a time of considerable fat deposition, particularly fat that is more centrally distributed, at least in men (38). Although the predominant predictor of fat gain in men at this time was age, appropriate participation in physical activity may modulate the overall effect (38). Thus late adolescence/young adulthood could be an appropriate time for targeted lifestyle interventions to limit fat deposition.

ACKNOWLEDGMENTS

This research was supported by funding from Health Research Council, the University of Otago, the Child Health Research Foundation, Otago Medical Research Foundation, and the Dunedin School of Medicine Dean's Bequest Fund.

DISCLOSURE

The authors declared no conflict of interest.

© 2009 The Obesity Society

REFERENCES

1. Kissebah AH, Krakower GR. Regional adiposity and morbidity. *Physiol Rev* 1994;74:761–811.
2. Taylor RW, Gold E, Manning P, Goulding A. Gender differences in body fat content are present well before puberty. *Int J Obes Relat Metab Disord* 1997;21:1082–1084.
3. Arfai K, Pitukcheewanont PD, Goran MI *et al.* Bone, muscle, and fat: sex-related differences in prepubertal children. *Radiology* 2002;224:338–344.
4. Garnett SP, Höglér W, Blades B *et al.* Relation between hormones and body composition, including bone, in prepubertal children. *Am J Clin Nutr* 2004;80:966–972.
5. Malina RM. Regional body composition: age, sex and ethnic variation. In: Roche AF, Heymsfield SB, Lohman TG (eds). *Human Body Composition*. Human Kinetics: Champaign, IL, 1996, pp 217–255.
6. Heude B, Kettaneh A, de Lauzon Guillaumin B *et al.*; Fleurbaix Laventie Ville Santé Group. Growth curves of anthropometric indices in a general population of French children and comparison with reference data. *Eur J Clin Nutr* 2006;60:1430–1436.
7. He Q, Horlick M, Thornton J *et al.* Sex and race differences in fat distribution among Asian, African-American, and Caucasian prepubertal children. *J Clin Endocrinol Metab* 2002;87:2164–2170.
8. He Q, Horlick M, Thornton J *et al.* Sex-specific fat distribution is not linear across pubertal groups in a multiethnic study. *Obes Res* 2004;12:725–733.
9. Dencker M, Thorsson O, Lindén C *et al.* BMI and objectively measured body fat and body fat distribution in prepubertal children. *Clin Physiol Funct Imaging* 2007;27:12–16.
10. Wang H, Story RE, Venners SA *et al.* Patterns and interrelationships of body-fat measures among rural Chinese children aged 6 to 18 years. *Pediatrics* 2007;120:e94–e101.
11. Ogle GD, Allen JR, Humphries IR *et al.* Body-composition assessment by dual-energy x-ray absorptiometry in subjects aged 4–26 y. *Am J Clin Nutr* 1995;61:746–753.
12. Huang TT, Johnson MS, Figueroa-Colon R, Dwyer JH, Goran MI. Growth of visceral fat, subcutaneous abdominal fat, and total body fat in children. *Obes Res* 2001;9:283–289.
13. Brambilla P, Bedogni G, Moreno LA *et al.* Crossvalidation of anthropometry against magnetic resonance imaging for the assessment of visceral and subcutaneous adipose tissue in children. *Int J Obes (Lond)* 2006;30:23–30.
14. Wells JC. A critique of the expression of paediatric body composition data. *Arch Dis Child* 2001;85:67–72.
15. Wells JC. Sexual dimorphism of body composition. *Best Pract Res Clin Endocrinol Metab* 2007;21:415–430.
16. Goulding A, Jones IE, Taylor RW, Manning PJ, Williams SM. More broken bones: a 4-year double cohort study of young girls with and without distal forearm fractures. *J Bone Miner Res* 2000;15:2011–2018.
17. Taylor RW, Jones IE, Williams SM, Goulding A. Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy X-ray absorptiometry, in children aged 3–19 y. *Am J Clin Nutr* 2000;72:490–495.
18. Goulding A, Jones IE, Taylor RW, Williams SM, Manning PJ. Bone mineral density and body composition in boys with distal forearm fractures: a dual-energy x-ray absorptiometry study. *J Pediatr* 2001;139:509–515.
19. McAuley KA, Williams SM, Mann JI *et al.* Intensive lifestyle changes are necessary to improve insulin sensitivity: a randomized controlled trial. *Diabetes Care* 2002;25:445–452.
20. Miller JC, Grant AM, Drummond BF *et al.* DXA measurements confirm that parental perceptions of elevated adiposity in young children are poor. *Obesity (Silver Spring)* 2007;15:165–171.
21. Sutherland TJ, Goulding A, Grant AM *et al.* The effect of adiposity measured by dual-energy X-ray absorptiometry on lung function. *Eur Respir J* 2008;32:85–91.
22. Taylor RW, Murdoch L, Carter P *et al.* Longitudinal study of physical activity and inactivity in preschoolers: the FLAME study. *Med Sci Sports Exerc* 2009;41:96–102.
23. Duke PM, Litt IF, Gross RT. Adolescents' self-assessment of sexual maturation. *Pediatrics* 1980;66:918–920.
24. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 2000;320:1240–1243.
25. Taylor RW, Keil D, Gold EJ, Williams SM, Goulding A. Body mass index, waist girth, and waist-to-hip ratio as indexes of total and regional adiposity in

- women: evaluation using receiver operating characteristic curves. *Am J Clin Nutr* 1998;67:44–49.
26. Webster-Gandy J, Warren J, Henry CJ. Sexual dimorphism in fat patterning in a sample of 5 to 7-year-old children in Oxford. *Int J Food Sci Nutr* 2003;54:467–471.
 27. Marini E, Cabras S, Rebato E *et al*. Sex differences in skinfold variability across human populations and during the life cycle. *Ann Hum Biol* 2007;34:377–392.
 28. Mast M, Körtzinger I, König E, Müller MJ. Gender differences in fat mass of 5-7-year old children. *Int J Obes Relat Metab Disord* 1998;22:878–884.
 29. Kirkby J, Metcalf BS, Jeffery AN *et al*. Sex differences in resting energy expenditure and their relation to insulin resistance in children (EarlyBird 13). *Am J Clin Nutr* 2004;80:430–435.
 30. Cowell CT, Briody J, Lloyd-Jones S *et al*. Fat distribution in children and adolescents—the influence of sex and hormones. *Horm Res* 1997;48(Suppl 5):93–100.
 31. Allison DB, Paultre F, Goran MI, Poehlman ET, Heymsfield SB. Statistical considerations regarding the use of ratios to adjust data. *Int J Obes Relat Metab Disord* 1995;19:644–652.
 32. Cole TJ, Fewtrell MS, Prentice A. The fallacy of using percentage body fat as a measure of adiposity. *Am J Clin Nutr* 2008;87:1959; author reply 1959–1960.
 33. Carey DG, Jenkins AB, Campbell LV, Freund J, Chisholm DJ. Abdominal fat and insulin resistance in normal and overweight women: direct measurements reveal a strong relationship in subjects at both low and high risk of NIDDM. *Diabetes* 1996;45:633–638.
 34. Goran MI, Gower BA. Relation between visceral fat and disease risk in children and adolescents. *Am J Clin Nutr* 1999;70:149S–156S.
 35. Owens S, Gutin B, Ferguson M *et al*. Visceral adipose tissue and cardiovascular risk factors in obese children. *J Pediatr* 1998;133:41–45.
 36. Daniels SR, Morrison JA, Sprecher DL, Khoury P, Kimball TR. Association of body fat distribution and cardiovascular risk factors in children and adolescents. *Circulation* 1999;99:541–545.
 37. Fox K, Peters D, Armstrong N, Sharpe P, Bell M. Abdominal fat deposition in 11-year-old children. *Int J Obes Relat Metab Disord* 1993;17:11–16.
 38. Nordström A, Neovius MG, Rössner S, Nordström P. Postpubertal development of total and abdominal percentage body fat: an 8-year longitudinal study. *Obesity (Silver Spring)* 2008;16:2342–2347.
 39. Matsudo SMM, Matsudo VKR. Self-assessment and physician assessment of sexual maturation in Brazilian boys and girls: concordance and reproducibility. *Am J Hum Biol* 1994;6:451–455.
 40. Roemmich JN, Clark PA, Mai V *et al*. Alterations in growth and body composition during puberty: III. Influence of maturation, gender, body composition, fat distribution, aerobic fitness, and energy expenditure on nocturnal growth hormone release. *J Clin Endocrinol Metab* 1998;83:1440–1447.
 41. Rosenbaum M, Leibel RL. Clinical review 107: role of gonadal steroids in the sexual dimorphisms in body composition and circulating concentrations of leptin. *J Clin Endocrinol Metab* 1999;84:1784–1789.
 42. Lovejoy JC, Sainsbury A; Stock Conference 2008 Working Group. Sex differences in obesity and the regulation of energy homeostasis. *Obes Rev* 2009;10:154–167.

Gender Differences Across Age in Motor Performance: A Meta-Analysis

Jerry R. Thomas
Departments of Physical Education
and Psychology
Louisiana State University

Karen E. French
Department of Physical Education
Louisiana State University

A meta-analysis was conducted to examine gender differences in motor performance during childhood and adolescence. Data were 64 studies yielding 702 effect sizes based on 31,444 subjects. Age was regressed on effect size, and the relation was significant for 12 of 20 tasks. Several types of age-related curves were found; the curve for a throwing task was the most distinctive. Five of the tasks followed a typical curve of gender differences across age. For eight tasks, gender differences were not related to age, and effect sizes were small. Results are discussed relating the development of gender differences to biological and environmental sources.

Across the childhood and adolescent years, gender differences have been reported in performance for many motor tasks. A typical description is one in which female and male performance differences are slight but favor males in early childhood. Performance rapidly accelerates linearly across childhood, with boys maintaining a slight but increasing advantage. At puberty, female performance levels off, whereas male performance continues to improve and may even accelerate.

How large are gender differences in motor performance? Can their rate of development be estimated from literature that are basically cross-sectional in nature? If a description of one or more gender difference curves across age is developed, can the sources of the differences be inferred from descriptive data? The purpose of this article is to evaluate by meta-analysis the gender differences in motor performance across childhood and adolescence and to suggest possible sources of the differences. The obvious potential sources of explanation are biology, environment, and their interaction.

Anastasi (1981) indicated that to progress from merely describing sex differences to explaining them the function of heredity and en-

vironment (nature-nurture, biology-culture) had to be considered. For example, she suggested that the biological acceleration of girls is the mechanism to explain girls' more rapid acquisition of language skills. Anastasi noted girls' physical acceleration, but seemed to imply that boys' advantage in gross motor skills during infancy and childhood has some biological basis (p. 198). Why would girls not perform gross motor skills more effectively than boys if they are physically and psychologically accelerated? Given the small size of the prepubertal gender differences in most motor performance tasks, are environmental factors a more likely source?

Why are Gender Differences in Motor Performance Present?

Biology

The physical characteristics of boys and girls are very similar prior to puberty. In fact, gender sameness rather than difference is a more appropriate descriptor of biological characteristics such as body type, body composition, strength, and limb lengths (Malina, 1984). Thus, biology seems to offer little explanation for motor performance differences prior to puberty.

Girls have their peak growth spurt approximately 2 years earlier than boys. Ultimately, this results in the termination of long bone growth earlier, which causes girls, on the average, to be shorter than boys (Espenschade &

Karen E. French is now at the Department of Physical Education, University of South Carolina, Columbia.

Requests for reprints should be sent to Jerry R. Thomas, School of Health, Physical Education, Recreation, and Dance, Louisiana State University, Baton Rouge, Louisiana 70803.

Eckert, 1980). Also, boys during and after puberty produce increasing amounts of testosterone, which is closely associated with increased muscle tissue. The ratio of muscle to fat is similar for boys and girls prepuberty. However, after puberty, this ratio remains approximately the same for females but doubles for males (Malina & Johnson, 1967). On the average, boys become taller and heavier than girls after puberty (National Center for Health Statistics, 1977). Boys have more lean body mass and less fat (Burmeister, 1965), greater arm and calf circumference (Roche & Malina, 1983), broader shoulders and narrower hips (Roche & Malina, 1983), and greater midarm muscle circumference and smaller triceps skinfold (Frisancho, 1981). Thus, in any motor task for which size and strength are an advantage, adolescent boys will have a biological advantage in performance when compared with adolescent girls because boys are larger and have more muscle.

Environment

An important potential source of environmental influence on gender differences in motor performance is the child's perception, which evolves over time, of the appropriate gender role (Greendorfer, 1980). In particular, the child's family, peers, teachers, and coaches are potential sources for learning a gender role regarding motor skill performance. The process of gender-role identification has been attributed to three sources: imitation, socialization, and self-socialization (Maccoby & Jacklin, 1974).

Several studies have reported that, during preschool years, both parents tend to emphasize the development of gross motor behavior in boys more than girls (for a brief review, see Maccoby & Jacklin, 1974, pp. 307-311). This included the fathers engaging in more rough play with boys and treating girls as more fragile. In particular, both parents (but especially fathers) reacted more negatively when boys chose to play with dolls than when girls chose "rough and tumble" games (Fling & Manosevitz, 1972; Lansky, 1967), a fact which seems pertinent to the development of gender differences in motor skills during early childhood. This early gender difference in motor perfor-

mance may also be influenced by parents' subtle messages that gross motor activities and some types of toys are more appropriate for boys than girls (Fagot, 1978). Thus, the gender differences in motor performance that are present as children enter elementary school may be largely socialized by parents, either by subtle coercion or by the children modeling what is perceived as sex-appropriate behavior.

If Sherif and Rattray (1976) are correct, physical education teachers and coaches in organized sport programs have treated the gender differences in motor performance exhibited in early childhood as naturally evolving biological factors. Even professional organizations such as the American Alliance for Health, Physical Education, Recreation and Dance (AAPHER) may contribute to gender differences by publishing separate norms for elementary school boys and girls for their physical fitness tests (AAHPER, 1976). This may lead to different expectations for boys and girls by teachers, coaches, parents, and the children themselves. As previously suggested, these motor performance differences are generally not large by the time children enter kindergarten or first grade and are most likely created by social factors. The treatment of differences as natural and biological may serve to increase them in many motor performance tasks during the elementary school years. Greendorfer (1980) and Housner (1981) provided interesting reviews and speculations about the role of physical educators and coaches in the development of gender differences in motor performance.

At puberty, gender differences in motor performance appear to be influenced by both biological and environmental factors. Although we must acknowledge the importance of biological changes as being closely associated with increasingly larger gender differences in many motor performance tasks, environmental factors may assume even greater importance at and following puberty. Boys are expected to be more masculine and girls more feminine, whereas in prepuberty some "tomboy" type behavior may have been socially acceptable in girls (Maccoby & Jacklin, 1974). Because of social pressures to conform, girls may be less inclined to participate in athletic activities and less motivated to perform well on motor tasks they do attempt. Thus, true gender differences may be overestimated.

A summary suggests that prior to puberty most gender differences in motor performance are socially induced by parents, peers, teachers, and coaches, although differences are by no means uniform and may include some type of gender-related predisposition toward certain motor tasks (a view originally suggested by Maccoby & Jacklin, 1974, pp. 363–364, as a framework to evaluate sex differences in psychological factors). Even though environmental pressures may be greater after puberty, biology plays an important role in the development of gender differences in many motor performance tasks, specifically those for which size and force production are important.

Gender Difference Inferences

We believe that quantifying the gender differences across age levels from early childhood to late adolescence provides a basis for theorizing about environmental and biological causes of any observed differences. If gender differences are small (even though reliable) across early childhood but begin to increase during the elementary school years, environment seems the more likely cause. This is particularly true if a remedial program eliminates most of the gender differences. The basis for this explanation is the previously presented evidence that parents and peers influence the early development of motor skills around sex-role models. These differences may be viewed as natural by teachers and coaches, who continue to contribute to their gradual increase during the elementary school years.

Biological factors seem to be implicated if large gender differences are noted during early childhood, particularly if these differences can be corroborated by cross-cultural findings and are difficult to reduce by training. However, even differences in biology seem to be subject to environmental reinforcement and enhancement. This is evident when a large difference noted in early childhood continues to increase during the elementary school years.

The influence of biology may also be evident in many motor tasks at puberty when male performance accelerates while female performance levels off, especially if task performance is enhanced by increased strength and size. But more likely a biology–environment interaction

is involved, as social pressures are intense to conform to the feminine or masculine sex role.

“For obvious ethical and practical reasons, research on human subjects cannot expose individuals to drastic and long-lasting variations in living conditions” (Anastasi, 1981, p. 188). Thus, we are forced to use relations found in large bodies of developmental data to infer cause–effect. This methodology is essentially the same as cross-cultural research; the source of the variation is simply vertical (across time) rather than horizontal (over cultures). However, the reader should be aware that these data are only correlational, making cause–effect conclusions tenuous.

Use of Meta-Analysis

Meta-analysis (Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985) offers an objective way to evaluate the large body of literature in which gender comparisons of motor performance are reported during childhood and adolescence. In many studies, motor performance may be measured in slightly different ways. For example, running speed may be tested over varying distances. If we express male–female differences in a standard metric (standard deviation units) effect size, these studies may be combined quantitatively in a meaningful way. Thus, running speed and its development can be analyzed as a concept rather than having to consider the individual measures of running speed.

Effect size provides a way of judging the size and meaningfulness of gender differences in motor performance. Because effect size is in standard deviation units, the size of the differences and degree of overlap between the distributions can be estimated. Cohen (1969) provided a means by which to judge effect sizes: 0.20 = small, 0.50 = medium, and 0.80 = large.

The use of meta-analysis eliminates two methodological problems. First, meta-analysis overcomes the problem (Jacklin, 1981) of doing regressions within gender and then inferring between-gender differences when predictors are entered in one gender’s prediction equation but not the other. The differences between genders are directly calculated as effect sizes prior to regressing age and study char-

acteristics on them. Second, meta-analysis allows the quantitative integration of a large number of cross-sectional and longitudinal studies of gender differences in motor performance. We believe that having a large number of effect sizes of gender differences in motor performance at a specific age provides the best available estimate for any true differences that may exist.

Thus, we believe that this meta-analysis of age-related changes in gender differences in motor performance provides the means to integrate and describe a large data set. In addition, it has considerable potential to further define theory about the development of sex differences, as well as to overcome some of the related methodological issues. In the following section, we have provided the details the reader needs for evaluating our work.

Method

Selection of the Data

Data were from studies that reported gender differences on motor performance during childhood and adolescence. Motor performance is defined as the outcome of movement (e.g., how fast a subject runs, how far a subject throws). In this study, motor performance was delimited to include (a) fundamental skills such as throwing, catching, running, striking, and jumping; (b) basic abilities such as balance and fine eye-motor coordination; (c) motor fitness items such as agility, arm strength, grip strength, flexibility, shuttle run, and sit-ups; and (d) information-processing responses such as reaction time, pursuit rotor tracking, and anticipation timing. The variables selected for study are those generally reported in studies and summaries in the motor-development literature (e.g., Corbin, 1980; Ridenour, 1978; Thomas, 1984; Wickstrom, 1983). Specifically excluded from the motor performance definition were cardiovascular fitness measures such as distance runs, step tests, and laboratory tests (treadmill walking and cycle ergometer riding).

Within the constraints of this definition of motor performance, a literature search was conducted using two computer bases—Educational Resources Information Center (ERIC) and PsycINFO. Health, Physical Education, and Recreation (HPER) Microform Publications (dissertations and theses) were searched by hand, as were eight journals identified as being likely to contain appropriate research reports: *Child Development*, *Developmental Psychology*, *Journal of Experimental Child Psychology*, *Journal of Human Movement Studies*, *Journal of Motor Behavior*, *Journal of Sport Psychology*, *Perceptual and Motor Skills*, and *Research Quarterly for Exercise and Sport*. The first study located was conducted in 1899, and the literature was searched through 1983.

From this search 176 studies were identified for initial consideration. Of these, 40 (23%) were from unpublished sources, mostly theses and dissertations. Only 64 (36%) of

the studies could be included in the meta-analysis. (The Appendix provides citations for each of these articles.) Of the 112 (64%) studies excluded (a list can be obtained from the first author), 32 (18%) had collapsed the design and data across age levels of more than 3 years, 45 (26%) had used both girls and boys but had not provided specific information about gender ratios, 31 (18%) did not provide the minimal information necessary to calculate effect size, and 4 (2%) were eliminated because the tasks were dropped. To be included a study had to provide male–female comparisons within the childhood and adolescent years.

Data Coded From Each Study

Effect sizes (ESs) were calculated for all male–female comparisons at each age level for motor performance tasks reported. An ES was obtained for a given task at a specific age by subtracting the mean for girls from the mean for boys and dividing by the standard deviation (Glass et al., 1981; Hedges & Olkin, 1985). When the mean for boys represented better performance, the ES was positive. When the girls averaged a better performance, the ES was negative.

To calculate ES, a pooled estimate of the standard deviation (weighted for sample size) was used (Hedges & Olkin, 1985). When means and standard deviations were provided, these techniques were used to estimate ES. When complete data are not available, ES can be estimated given some minimal statistical information. These procedures were applied to estimate ES when possible (Glass et al., 1981; Hedges & Olkin, 1985).

A common practice in gender difference research is to collapse the design across gender if the test of gender is not significant. Frequently, when this is done neither the means and standard deviations nor *F* ratios are provided. The author could be contacted and asked to provide the information, but this technique has proved unsuccessful in the past (Hyde, 1981). The ES of interest could be discarded, but this tends to bias the data to studies reporting significant gender differences. Finally, an ES as 0 could be used because by definition the statistical test is evaluating the null hypothesis of no reliable difference between the means. The data reported are with 0 ES included because the difference with and without was trivial.

Some studies included boys and girls in the sample but did not test the gender effect or report means and standard deviations by gender. These articles were excluded because there is no way to estimate an ES.

Effect size is still a biased estimator when sample sizes are small. Hedges (1981) and Hedges and Olkin (1985) provided a correction factor for small ESs. This factor— $c = 1 - 3/(4(n^m + n^f - 2) - 1)$, where n^m = number of boys and n^f = number of girls—is multiplied by the ES, thereby correcting overestimation of ES with a small sample size. All ESs (now labeled ES') in this study were corrected for bias using this formula.

In addition to ES', omega-squared (ω^2) was calculated for each boy–girl comparison using the formula provided by Tolson (1980). The use of ω^2 allows an estimate of the percent of variance accounted for by gender in the dependent measure. Or ω^2 estimates how well a specific motor performance task can be predicted if the gender of the subject is known. Thomas and Nelson (1985) discussed the need for considering the magnitude of the difference

(ω^2) as well as the reliability (significance) of the difference. Our major interest was to relate the ES' for gender differences in 20 motor performance tasks to age. However, five additional characteristics of each study were coded, and their relation to ES' was calculated.

1. *Internal validity* was coded as high, average, or low, using qualities of the study such as representativeness of the sample, quality of the motor performance test selected, and appropriateness of the measurement schedule.

2. *Gender of the first author* was used, based on literature suggesting that the gender of the experimenter is sometimes related to gender differences in motor performance (Rikli, 1976). This assumes that the first author is the experimenter, an assumption that may be incorrect in some instances. Thus, any relations found should be treated with caution.

3. *Type of manuscript* refers to published versus unpublished articles. Theses/dissertations were considered unpublished. If a thesis/dissertation was later published, it was coded as published and eliminated from the unpublished list.

4. *Year published* was coded as studies published before 1970 or 1970 and later. This date was selected because considerable change in sex roles occurred in the 1960s and 1970s (Anastasi, 1981, p. 202). More specifically, Title IX was passed in 1972. Thus, we wanted to see if changes in sex roles and opportunities were reflected in gender differences in motor performance.

5. *Number of male and female subjects in each study* was coded.

Characteristics of the Data

Sixty-four studies were included in the analysis. Several of the studies tested the subjects on more than one motor performance task: 16 studies had data on each subject for more than three motor tasks; 8 had three tasks; 14 had two tasks; and 26 had only one task. To some extent, when more than one measure is taken on each subject, task performance is correlated. However, these correlations are generally low for the tasks included, so we treated the tasks as if they were independent. We believe this is a valid assumption. The correlation of tests of the same characteristic (e.g., balance) are generally low (Johnson & Nelson, 1979). The specificity hypothesis (Henry, 1968) indicates that motor abilities are task specific, and that two similar tasks (e.g., throwing a football and a baseball) tend to have a correlation of zero. This is because the abilities underlying these tasks are different. Thus, correlations among different motor performance tasks (e.g., throwing and running) are expected to be low, and evidence generally supports this hypothesis (Schmidt, 1982, p. 401).

The 64 studies included 31,444 subjects of which 15,926 were boys and 15,518 were girls. Of the 64 studies included, 24 used fewer than 100 subjects, 33 used between 101 and 600 subjects, 6 used between 601 and 1,000 subjects, and 1 used more than 9,000 subjects. Several studies measured subjects on more than one motor performance task. Table 1 shows the number of measurements made for boys and girls by age level for the 64 studies.

Combining Effect Sizes

Hedges (1981) and Hedges and Olkin (1985) showed that, because of the way ES' is distributed, the variance

(s^2) of each ES' can be calculated directly, and they provided the formula. From this formula, the observation can be made that the accuracy of ES' is a function of sample size and ES'. Because the accuracy varies with this function, each ES' should be weighted by the reciprocal of its variance prior to combining ES' (Hedges, 1982; Hedges & Olkin, 1983, 1985). Thus, studies based on larger samples should be more precise estimators of true gender differences and, therefore, are given greater weight in the averaging of ES'. An overall or weighted mean estimate can be obtained by the formula provided by Hedges and Olkin (1985). An estimate of the variance for the weighted mean ES' is obtained from the bottom half of the same formula. These formulas were used in combining and estimating the variance of the ES'.

Results

We calculated the weighted mean of ES' for each gender comparison for all tasks. Then we examined the number of ES's for each task for all studies. When the number of studies was less than three, the task was eliminated, with one exception: Several tasks—hole punching, peg shifting, manual dexterity, tracing, and turning a screw (called speed prehension)—were combined and called fine eye-hand coordination. Table 2 lists those tasks that met the criteria for retention and those that were eliminated. Table 3 contains the number of measurements made on male and female subjects for each of the 20 tasks that were retained for the analysis.

Table 1
Number of Measurements for Boys and Girls by Age Level Included in the Meta-Analysis

Age (years)	No. for boys	No. for girls	Total
3	316	259	575
4	863	805	1,668
5	1,855	1,816	3,671
6	1,980	1,733	3,713
7	6,325	5,808	12,133
8	3,378	3,578	6,956
9	2,374	2,332	4,706
10	4,256	4,054	8,310
11	6,371	6,012	12,383
12	4,696	4,892	9,588
13	5,221	5,447	10,668
14	3,021	3,108	6,129
15	3,369	3,112	6,481
16	3,256	3,162	6,418
17	3,647	2,865	6,512
20	142	142	284
Totals	51,070	49,125	100,195

Table 2
Motor Performance Tasks Included and Excluded

Task	No. studies	No. ES's	Task	No. studies	No. ES's
Included			Included (continued)		
Agility	9	19	Throw velocity	5	13
Anticipation timing	3	23	Vertical jump	5	20
Arm hang	3	16	Wall volley	4	32
Balance	14	71	Excluded		
Catching	4	25	Choice RT	2	10
Dash	19	82	Dribble	1	2
Fine eye-hand	5	30	8 choice RT	2	21
Flexibility	5	13	4 choice RT	2	21
Grip strength	4	42	Free throw	1	4
Long jump	19	85	Hurdle jump	2	7
Pursuit rotor	5	14	Leg torque	1	5
Reaction time	6	42	Movement time	2	11
Shuttle run	7	33	Pull-ups	2	13
Sit-ups	7	36	Push-ups	2	7
Tapping	4	34	Squat thrust	1	8
Throw accuracy	5	14	Striking	2	15
Throw distance	11	58	2 choice RT	2	21

Note. RT = reaction time. Total number of studies included = 144; this is more than the total number of studies because some contained more than one task. Number of ES's for studies included = 702. Total number of studies excluded = 22. Total number of ES's for studies excluded = 145.

The first issue is to test if ES' has homogeneity within each task (Hedges & Olkin, 1985). This test (*H* statistic) is the total sum of squares tested as a χ^2 using the number of ES's minus 1 as the degree of freedom. This test was significant ($p < .05$) for 9 of the 20 tasks (balance, grip strength, shuttle run, throw for velocity, vertical jump, dash, long jump, sit-ups, and throw for distance). However, we hypothesized a priori that a major source of this lack of homogeneity would be the large age range in the data. Other possible sources of a lack of homogeneity are outliers and study characteristics. Therefore, we used a weighted regression as outlined by Hedges and Olkin (1983, 1985, chap. 8) to examine the correlation of ES' to age within each task. We allowed the fit of the regression line to be linear (age), quadratic (age²), or cubic (age³).

Tasks in Which Gender Differences in Motor Performance and Age Were Related

The first step was to identify which tasks had ES's for gender that were age related. Table 4 provides a summary of the regression statistics for the 12 tasks (out of 20) that correlated significantly with age. (The *H* statistic for these

12 tasks is the sum of the χ^2 for the regression plus the χ^2 for within tested against the sum of the degrees of freedom for each.)

Table 3
Number of Measurements by Gender for the 20 Motor Performance Tasks

Task	No. for boys	No. for girls	Total
Agility	1,869	1,643	3,512
Anticipation timing	303	433	736
Arm hang	268	234	502
Balance	2,269	2,122	4,391
Catching	742	698	1,440
Dash	8,191	7,830	16,021
Fine eye-hand	1,035	982	2,017
Flexibility	1,074	995	2,069
Grip strength	1,446	1,338	2,784
Long jump	9,406	8,953	18,359
Pursuit rotor	245	237	482
Reaction time	1,478	1,597	3,075
Shuttle run	5,953	5,714	11,667
Sit-ups	5,805	5,545	11,350
Tapping	759	706	1,465
Throw accuracy	688	624	1,312
Throw distance	7,754	7,558	15,312
Throw velocity	480	465	945
Vertical jump	384	423	807
Wall volley	921	1,028	1,949

Hedges and Olkin (1983, 1985) argued that the typical way of interpreting this regression is inappropriate. They indicated that the first step is to determine if the regression is significant by treating the sum of squares for the regression as a χ^2 with the degrees of freedom associated with the regression (model). Note in Table 4 that the column labeled χ^2_{reg} is this test with its associated degrees of freedom.

Once the regression is declared significant, a test should be made to determine if the model is correctly specified. "The test for model specification provides a basis for deciding whether the variation in effect magnitude is accounted for by the explanatory variables in the model" (Hedges & Olkin, 1983, p. 137). Hedges and Olkin (1983, 1985) indicated that the sum of squares for within (error) is the appropriate test of model specification.

This sum of squares is treated as a χ^2 , with the degrees of freedom associated with the within factor (see Table 4, column labeled χ^2_{wi}). Note in Table 4 that we have placed the eight tasks (balance, catching, grip strength, pursuit rotor, shuttle run, tapping, throwing velocity, and vertical jump) where the model is correctly specified in the top part of the table and the four tasks (dash, long jump, sit-ups, and distance throw) where the model is incorrectly specified at the bottom of the table. The fact that the data do not fit the model means that the results should be viewed with caution.

As indicated by the data in Table 4, the line of best fit is linear (see column labeled Age where the top value is the beta and the bottom value is the z score of the beta, all corrected according to procedures indicated by Hedges & Olkin, 1985) for balance, pursuit rotor, tapping, throwing velocity, and throwing distance. The best fit was quadratic (column labeled Age²) for catching, grip strength, shuttle run, vertical jump, dash, long jump, and sit-ups.

For the figures that follow, the weighted $\bar{ES}$ is bounded by a 95% confidence interval for the 12 tasks in which gender differences are related to age. Clearly, throwing for either velocity or distance (Figure 1) is very different from the other 10 motor performance tasks. Boys exceed girls in throwing velocity by 1.5 standard deviation units as early as 4 to 7 years of age. This difference rapidly increases so that by 12 years of age (last age with sufficient data for a comparison) the boys' performance ex-

ceeds the girls' performance by 3.5 standard deviation units. Throwing for distance, which to a large extent is dependent on the velocity of the throw, follows a similar pattern. The boys exceed the girls by 1.5 standard deviation units as early as 2 to 4 years of age. The increase is linear through puberty to 17 years of age when male performance is 3 standard deviation units better than female performance. The weighted $\bar{ES}$ collapsed over age is 1.98 standard deviation units for throwing velocity and 2.18 for throwing distance. These data indicate that there is very little overlap in the distribution of throwing performance for boys and girls.

Balance, pursuit rotor tracking, and tapping (Figure 2) seem to have similar patterns. Basically, the gender difference is zero until puberty (10 to 13 years of age) when the male performance becomes better than female performance: Weighted $\bar{ES}$ for balance is about 1 standard deviation unit, about 0.75 for pursuit rotor tracking, and about 0.50 for tapping. However, when averaged across age, the weighted $\bar{ES}$ s are relatively small: 0.09 for balance, 0.11 for pursuit rotor tracking, and 0.13 for tapping.

Five of the remaining tasks follow what has become regarded as the typical pattern for a motor performance task across childhood and adolescence. The weighted $\bar{ES}$ s are slightly in favor of the boys in early childhood by 0.25 to 0.50 standard deviation units. This difference increases somewhat during middle childhood to 0.50 to 1.00 standard deviation units. The difference increases to between 1.00 and 2.00 standard deviation units after puberty. The tasks following this pattern are the dash and situps (Figure 3) and the long jump, grip strength, and shuttle run (Figure 4).

For the vertical jump (Figure 5) the weighted $\bar{ES}$ is essentially zero until puberty, when it increases to more than 1 standard deviation unit in favor of boys. Catching (Figure 5) follows a U-shaped pattern; boys perform about 0.75 standard deviation units better than girls during preschool. Then the differences become smaller, dropping to 0.25 units until postpuberty, when they again go up to 0.75 units. We have no data points after 13 years of age.

The model was not correctly specified for the shuttle run (Table 4) until three additional variables were included to reduce the within-

Table 4
Age-Related Studies: Regression Summary for ES'

Task	No. studies	No. ES'	$\overline{ES'}$	χ^2_{reg}	χ^2_{wi}	Age	Age ²	Validity of study	Year pub.	Gender of 1st author	Type article
Studies for which model is correctly specified											
Balance	14	67	0.09	42.9** 1 <i>df</i>	76.0 65 <i>df</i>	0.067** 6.55 <i>z</i>					
Catching	4	23	0.43	13.4** 2 <i>df</i>	15.6 20 <i>df</i>	-0.226* 1.96 <i>z</i>	0.019** 2.62 <i>z</i>				
Grip strength	4	37	0.66	78.4** 2 <i>df</i>	45.2 34 <i>df</i>	-0.419** 5.24 <i>z</i>	0.024** 6.43 <i>z</i>				
Pursuit rotor	5	14	0.11	5.0* 1 <i>df</i>	4.5 12 <i>df</i>	0.125* 2.23 <i>z</i>					
Shuttle run	7	28	0.32	673.4** 5 <i>df</i>	29.6 22 <i>df</i>	-0.002 0.034 <i>z</i>	0.006** 2.39 <i>z</i>	1.40** 14.22 <i>z</i>		-1.96** 9.40 <i>z</i>	1.91** 11.44 <i>z</i>
Tapping	4	34	0.13	11.0* 1 <i>df</i>	21.5 32 <i>df</i>	0.080** 3.31 <i>z</i>					
Throw velocity	5	12	2.18	42.2** 1 <i>df</i>	11.3 10 <i>df</i>	0.454** 6.49 <i>z</i>					
Vertical jump	5	20	0.18	62.3** 1 <i>df</i>	16.6 17 <i>df</i>	-0.607 1.95	0.042** 2.62				
Studies for which model is incorrectly specified											
Dash	19	66	0.63	1061** 3 <i>df</i>	103** 62 <i>df</i>	-0.298** 8.42 <i>z</i>	0.021** 2.91 <i>z</i>				0.383** 7.42 <i>z</i>
Long jump	19	68	0.54	1212** 2 <i>df</i>	123** 65 <i>df</i>	-0.337** 4.95 <i>z</i>	0.023** 16.45 <i>z</i>				
Sit-ups	7	29	0.64	697** 4 <i>df</i>	79** 24 <i>df</i>	0.799** 2.87 <i>z</i>	-0.022** 4.89 <i>z</i>		-0.975** 10.0 <i>z</i>	-0.127** 5.80 <i>z</i>	
Throw distance	11	47	1.98	433** 1 <i>df</i>	87** 45 <i>df</i>	0.153** 20.8 <i>z</i>					

Note. ES = effect size. Year pub. = published before 1970 or not. Type article = published/unpublished article. *z* = *z* score.

* $p < .05$. ** $p < .01$.

sum of squares: the validity of the study (column labeled Validity in Table 4), gender of the first author, and type of article (published or unpublished). An interpretation of these characteristics is that larger weighted $\overline{ES}$'s for gender (better male performance) were associated

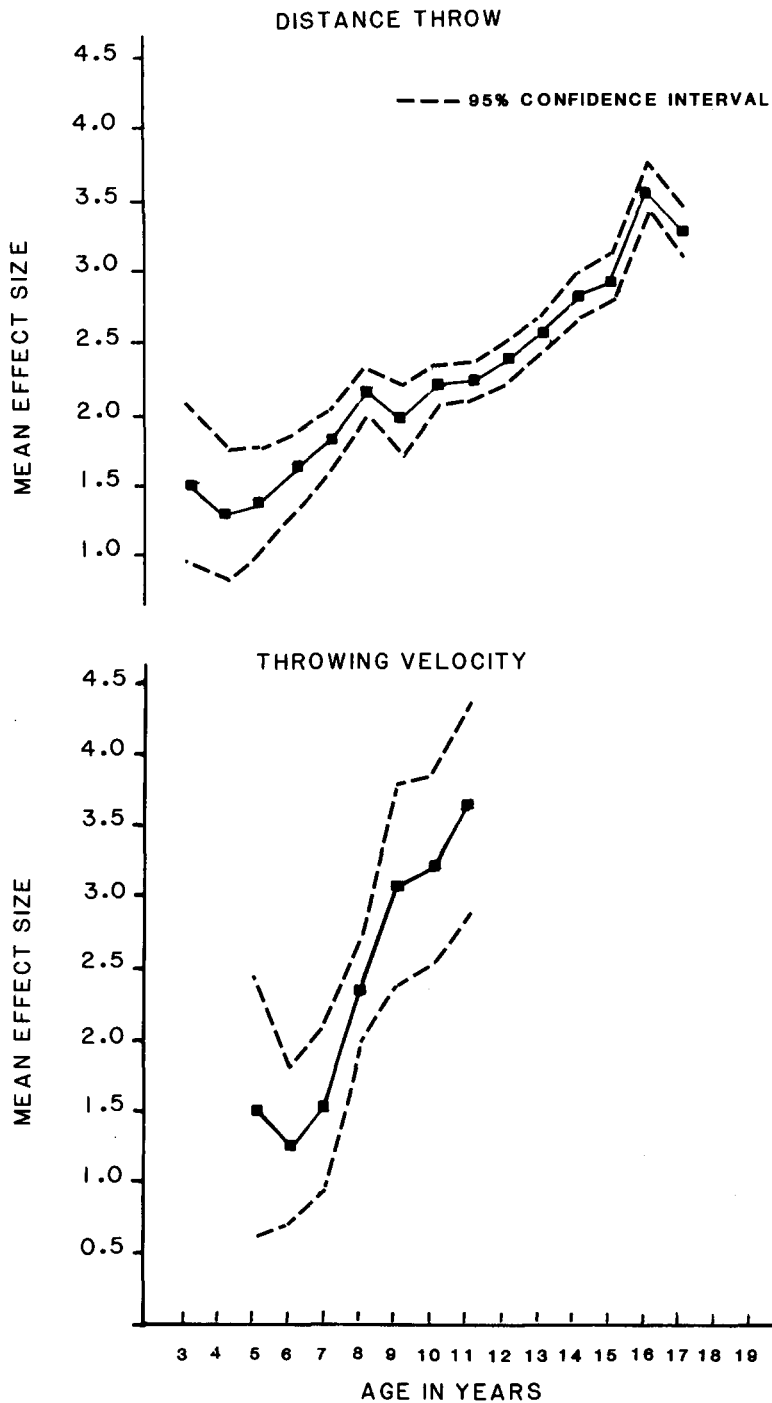


Figure 1. $\overline{ES}$ by age and gender for throwing for velocity and distance.

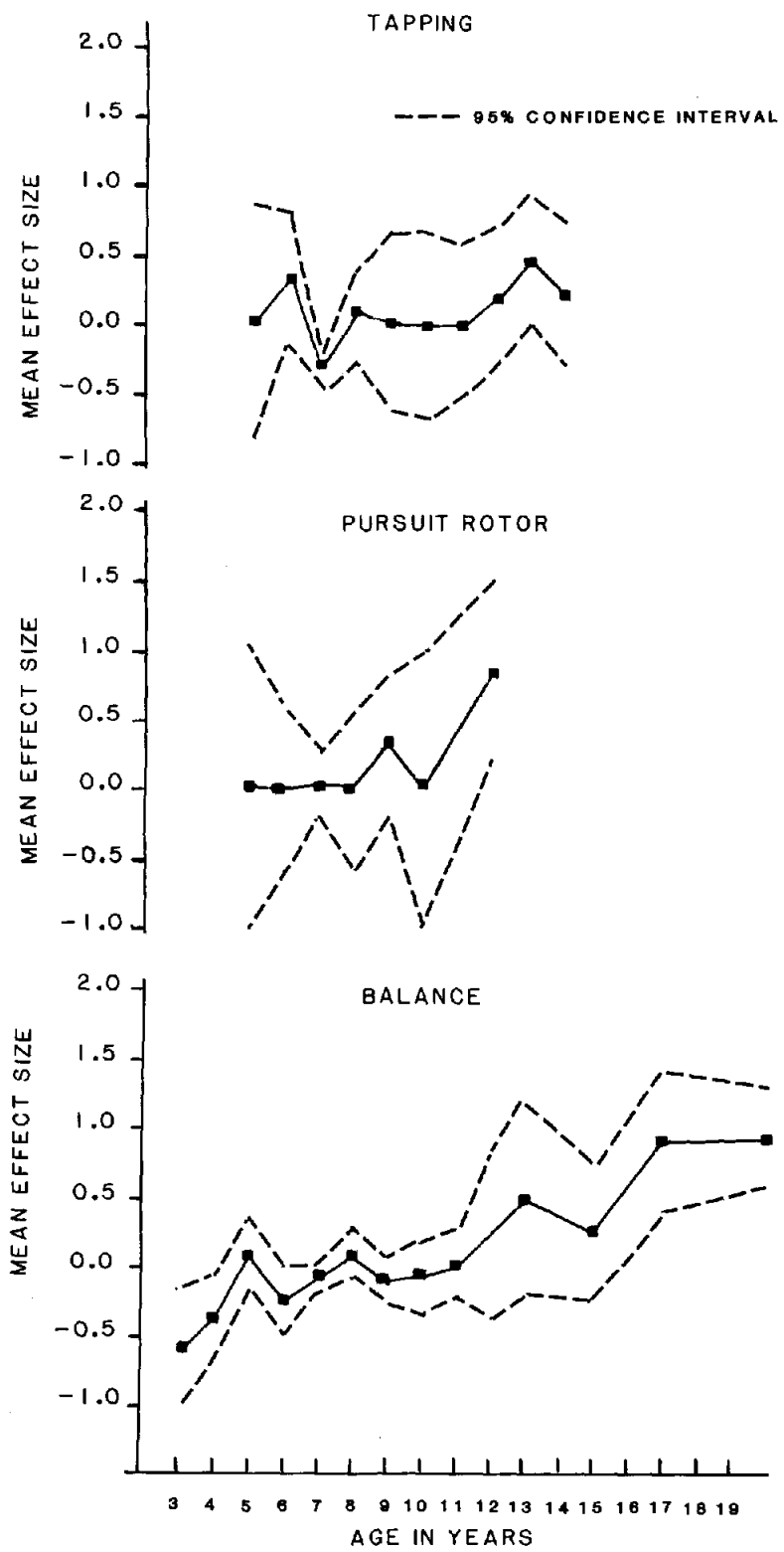


Figure 2. $\overline{ES}$ by age and gender for balance, pursuit rotor tracking, and tapping.

with studies published since 1970, having higher internal validity, and women as first author. For the tasks in which the model was not correctly specified, a significant amount of

variance was accounted for by some of the coded characteristics in addition to age. Sit-ups had larger weighted $\overline{ES}$'s associated with studies published before 1970 and studies in which women were first author. The dash had

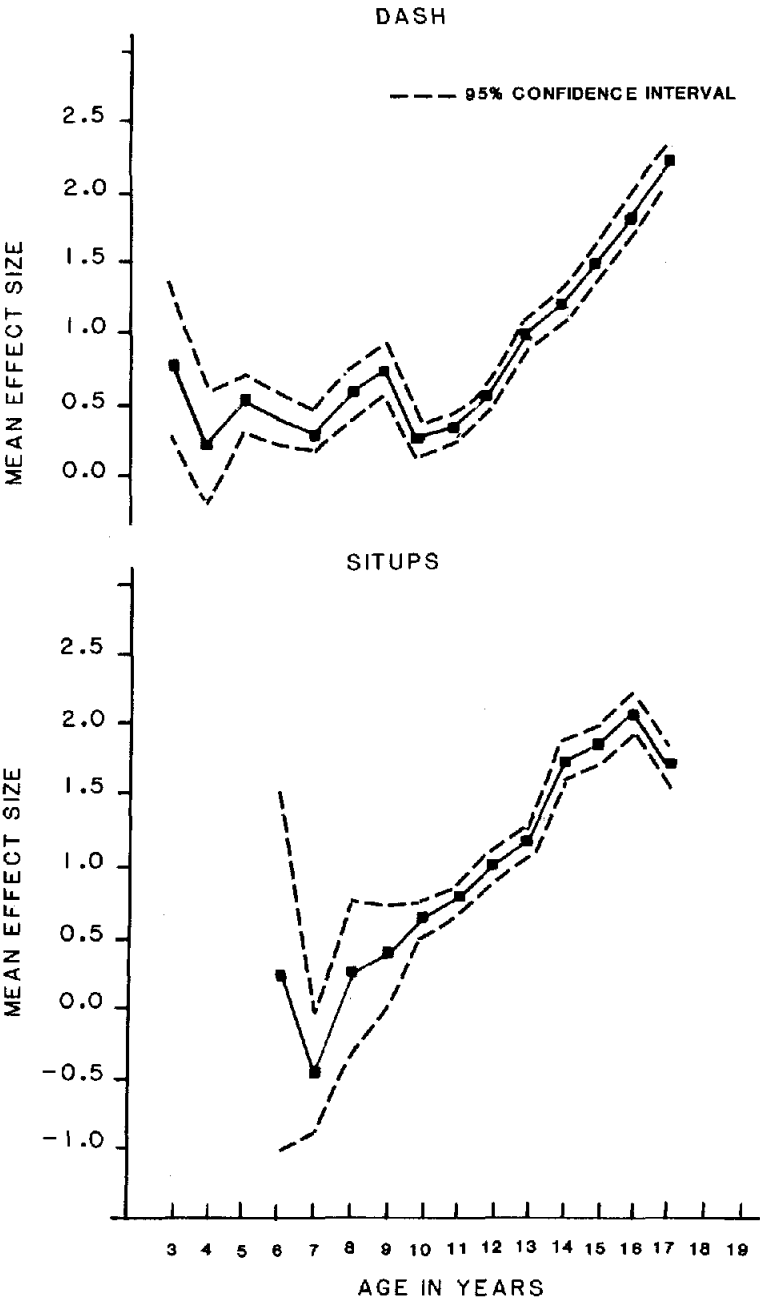


Figure 3. $\overline{ES}$ by age and gender for dash and sit-ups.

larger weighted $\overline{ES}$'s associated with published as opposed to unpublished articles.

In any data set, outliers (extreme and unrepresentative values) are likely to exist, even when data points are ES 's rather than individual performances. Outliers are particularly likely in data sets including young children (Thomas, 1984). An appropriate way to test for outliers in a meta-analysis using regression is to output the residuals from the regression, take the absolute value of the residual, calculate the mean and standard deviation of the residuals, and change the residuals to z scores (Hedges & Olkin, 1985). By definition (because the mean of a z score is 0 with a standard deviation of 1) any standardized residual larger than 2 is an outlier because it falls outside 95% of the score distribution. This procedure was used to identify outliers in the age-related data set. The data in Table 4 have the outliers eliminated. Table 5 includes the number and percentages of outliers eliminated for each task. An upper limit of eliminating 20% of the data was set and the criterion for declaring a data point an outlier was established as a z score of 2.

Although allowing 20% of the data to be declared as outliers may seem high, Table 5 shows that half of the tasks (6 of 12) actually had less than 10% of the ES 's eliminated as outliers, only 4 tasks had more than 15% of the ES 's eliminated, and only 1 task (long jump) actually reached the 20% maximum. Note that the four tasks (dash, long jump, sit-ups, and throw for distance) that had the highest percentage of outliers (19%–20%) were the same four for which the model was incorrectly specified (Table 4). This suggests that ES 's for gender for these four tasks are not derived from the same population of ES 's. Procedural differences among studies using the same general tasks may be the cause (e.g., dashes of varying lengths and sit-ups with time limits as compared with sit-ups with no time limits).

Hedges and Olkin (1985) believed the use of omega-squared (ω^2) as the dependent variable in a meta-analysis was inappropriate because the direction of the gender difference was not specified. However, the procedure is not without precedent (Hyde, 1981), and we believe analyzing ω^2 is useful because it represents the strength of the association between gender and the various motor performance tasks. Ta-

ble 6 has a summary of the results of a standard regression analysis of age on ω^2 . We have included the 12 tasks that were related to age in the earlier analysis of the weighted ES '. This table provides data for each task in which ω^2 's were significantly related to age in a linear (shuttle run, sit-ups, throwing distance, and throwing velocity) or quadratic (balance, catching, dash, grip strength, long jump, pursuit rotor, and vertical jump) manner, or in one instance (tapping) in which the task was not related to age. The ω^2 for gender accounted for the most variance (averaged across age levels) in the two throwing tasks: 51% in throwing velocity and 47% in throwing distance. In other words, at any age about one half of the variance in throwing performance between boys and girls can be estimated by knowing the performer's gender.

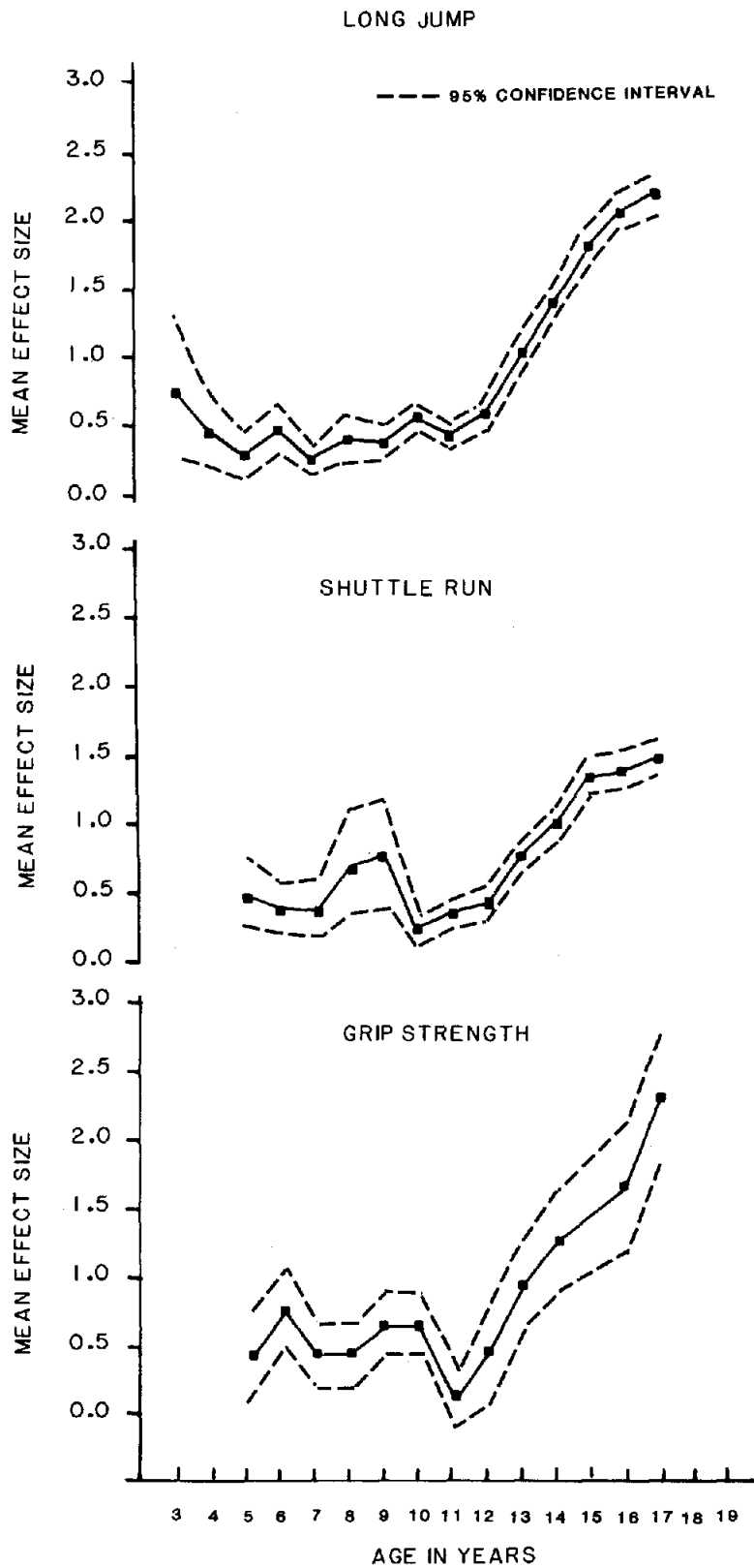
The developmental patterns for ω^2 are similar to those for ES '. Thus, we have not included them. For example, in balance, catching, dash, grip strength, long jump, sit-ups, and vertical jump, little variance ($\omega^2 < .10$) is accounted for until about puberty. At puberty, there is a rapid rise in variance accounted for with ω^2 between .20 and .65. Thus, knowing the gender of the subject postpuberty allows a reasonable prediction of performance, especially in speed (dash and shuttle run), strength and endurance (grip strength and sit-ups), and power (long jump and vertical jump) motor performance tasks. The shuttle run is slightly different from the other tasks, accounting for a great amount of variance prepuberty (10%–20%).

As we showed with ES ', throwing for either distance or velocity is clearly very different from the other types of motor performance tasks. Gender accounts for 35% to 60% of the variance before puberty and 60% to 70% after puberty. Given that gender is a dichotomous variable, the postpuberty prediction is approaching the upper limit of this relation. For tapping and pursuit rotor tracking, gender accounts for none of the variance prior to puberty and very little ($\omega^2 < .15$) after puberty.

Tasks for Which Gender Differences in Motor Performance Were Not Age Related

In Table 7 the summary data are reported for the motor performance tasks (8 of 20) for which gender differences were not related to

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



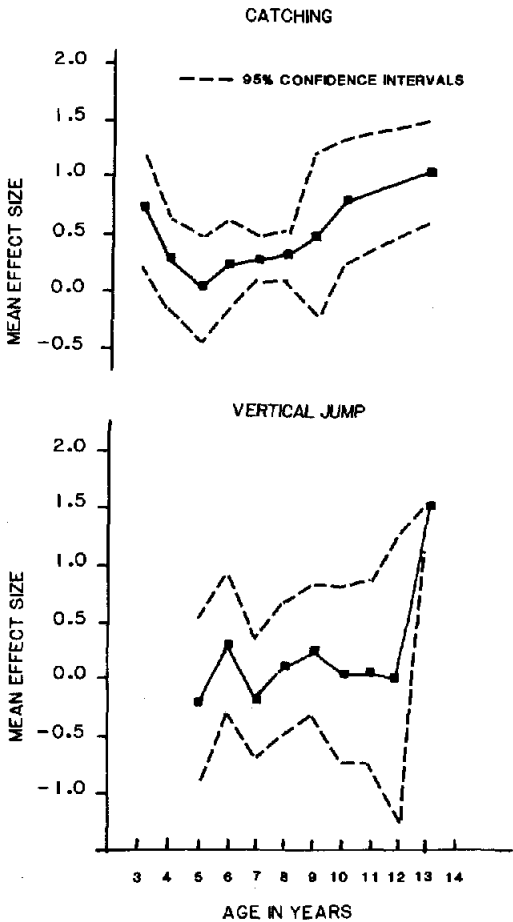


Figure 5. $\overline{ES}$ by age and gender for vertical jump and catching.

age. The homogeneity statistic (H) is tested as a χ^2 with degrees of freedom as the number of ES 's minus 1. The H was not significant ($p > .05$) for any of these tasks. The χ^2 test of the sum of squares for the regression was not significant ($p > .05$) for any of the 8 tasks.

The weighted $\overline{ES}$ was large for throwing accuracy (0.96) and the wall volley (0.83), both being close to 1 standard deviation unit. The percent of variance accounted for (ω^2) by these two tasks was also the largest of this group. The weighted $\overline{ES}$'s were small (0.18 to 0.38) for agility, anticipation timing, flexibility, reaction time, and fine eye-hand coordination. Note that flexibility ($\overline{ES}' = -0.29$) and fine eye-motor coordination ($\overline{ES}' = -0.21$) were

two tasks at which female performance was consistently better than male performance. The variance accounted for by gender in these five tasks was small ($\omega^2 < .08$). The arm hang showed no meaningful difference in $\overline{ES}'$ (0.01) or in variance accounted for ($\omega^2 = .01$).

Discussion

These data do not support the notion of uniform development of gender differences in motor performance across childhood and adolescence. Effect sizes were related to age in only 12 of the 20 (60%) motor performance tasks. These relations can be placed into several groups based on the shapes of the curves in Figures 1 to 5.

Tasks for Which Age and Gender Are Related

The major issue is how biology and environment influence the development of gender differences across childhood and adolescence. Can the nature of the developmental curves be used to make inferences about the causes of gender differences in motor performance?

Typical motor performance curve. If a typical motor performance curve exists (as described earlier) that reflects the development of gender differences, it is depicted in Figures 3 and 4 for the tasks of dash, sit-ups, long jump, grip strength, and shuttle run. In general these five tasks show small to moderate effect sizes (0.20 to 0.50) favoring boys in early childhood. The effect sizes remain moderate across the elementary school years for four of the tasks, but increase during elementary school for sit-ups. The girls, upon reaching puberty, appear slightly to reduce the difference on the dash, grip strength, and the shuttle run, but not in the long jump and sit-ups. When the boys reach puberty, all five tasks are influenced in the same way—they increase their advantage over girls rapidly until it is between 1.5 and 2 standard deviation units at 17 years of age. The nature of these five curves can be viewed with considerable trust because the 95% confidence intervals are tight.

Although girls on reaching puberty may close the performance difference on some of these five tasks, when boys reach puberty their

Figure 4. $\overline{ES}$ by age and gender for long jump, grip strength, and shuttle run.

Table 5
Number and Percentage of Outliers by Task

Task	No. studies	No. ES's	No. ES's lost as outliers	No. ES's kept	% of ES's lost
Balance	14	71	4	67	5.6
Catching	4	25	2	23	8.0
Dash	19	82	16	66	19.5
Grip strength	4	42	5	37	13.5
Long jump	19	85	17	68	20.0
Pursuit rotor	5	14	0	14	0.0
Shuttle run	7	33	5	28	15.0
Sit-ups	7	36	7	29	19.0
Tapping	4	34	0	34	0.0
Throw distance	11	58	11	47	19.0
Throw velocity	5	13	1	12	8.0
Vertical jump	5	20	0	20	0.0

increase in size and muscle tissue is dramatically reflected in better performance across all five tasks. The average performance of boys and girls of 2 standard deviation units apart at age 17 shows that there is little overlap between the two distributions. This is an expected finding for tasks in which speed, power, muscular strength, and endurance are important.

The differences prior to puberty are more moderate and, we believe, more likely to reflect environmental influences. When differences between the mean performances of boys and girls are less than 0.50 standard deviation units, many of the girls are performing better than many of the boys. If equal expectations, encouragement, and practice opportunities were provided by parents, teachers, and coaches,

differences of this size could probably be eliminated. Opportunities and encouragement to practice may be the key issue; Halverson, Robertson, and Langendorfer (1982) reported that seventh-grade boys remembered practicing throwing more over the years than did seventh-grade girls.

Several authors (Greendorfer, 1980; Housner, 1981; Sherif & Rattray, 1976) suggested that teachers and coaches take as biological the gender differences evident in motor performance when children begin elementary school. The implication is that these differences are maintained or increased because teachers and coaches have different expectations for boys than for girls. Although these data cannot speak directly to expectations, the differences

Table 6
Summary Data for Regression Using ω^2

Task	Age	Age ²	ω^2	R ²
Balance	$F(1, 64) = 22.24^{**}$	$F(1, 64) = 44.41^{**}$	.03	.63
Catching	$F(1, 20) = 12.37^{**}$	$F(1, 20) = 25.79^{**}$	.05	.81
Dash	$F(1, 63) = 38.01^{**}$	$F(1, 63) = 72.89^{**}$	.10	.79
Grip strength	$F(1, 34) = 26.48^{**}$	$F(1, 34) = 41.28^{**}$	.10	.74
Long jump	$F(1, 65) = 73.63^{**}$	$F(1, 65) = 127.19^{**}$	.08	.81
Pursuit rotor	ns	$F(1, 11) = 7.38^*$	.02	.70
Shuttle run	$F(1, 26) = 46.36^{**}$	ns	.12	.64
Sit-ups	$F(1, 27) = 22.93^{**}$	ns	.14	.46
Tapping	ns	ns	.01	.04
Throw distance	$F(1, 45) = 31.72^{**}$	ns	.47	.41
Throw velocity	$F(1, 10) = 30.89^{**}$	ns	.51	.75
Vertical jump	$F(1, 17) = 13.61^{**}$	$F(1, 17) = 17.47^{**}$	.03	.64

* $p < .05$. ** $p < .01$.

Table 7
Summary Data for Motor Performance Tasks Where Gender Differences Were Not Related to Age

Task	Age range (years)	No. ES's	No. studies	<i>H</i> ^a	$\overline{ES'}$	<i>s</i> ²	$\overline{\omega}^2$
Agility	3–17	19	9	27.15	0.21	.0019	.04
Anticipation timing	7–20	23	3	18.43	0.38	.0066	.07
Arm hang	3–12	16	3	7.86	0.01	.0100	.01
Fine eye–motor	3–10	30	5	35.82	–0.21	.0022	.04
Flexibility	5–10	13	5	19.01	–0.29	.0023	.02
Reaction time	5–20	42	6	38.90	0.18	.0013	.02
Throw accuracy	6–11	14	5	12.81	0.96	.0041	.24
Wall volley	7–13	32	4	21.31	0.83	.0023	.15

^a*H* statistic is treated as a χ^2 with *df* equal to the number of ES's – 1. None of these are significant at *p* < .05.

present in these five tasks when children begin elementary school are maintained or increased.

Other motor performance curves. A second group of tasks (Figure 2)—balance, pursuit rotor tracking, and tapping—is similar in nature. Gender differences are not present during the elementary school years but increase to a moderate level ($\overline{ES'} = 0.50$) favoring boys at about 11 to 12 years of age. In our opinion, this small an increase is more likely to reflect increasing environmental pressures rather than any biological factor related to puberty. These three tasks do not seem to be related to the strength, endurance, and size increases noted in boys at puberty. The finding is particularly interesting for balance because balance is a task in which girls have been reported as performing better than boys (Robertson, 1984). Although girls do have a slight advantage in early childhood (3 to 4 years of age), the gender difference is essentially zero until puberty, when male performance is moderately better.

The vertical jump (Figure 5) is similar to balance, pursuit rotor tracking, and tapping in that there are no gender differences until puberty. At puberty boys show a large increase in performance so that they are more than 1 standard deviation unit better than girls. The lack of prepuberty gender differences favoring boys is unexpected. We expected that vertical jump performance would be similar to performance in the long jump, dash, and the shuttle run as reported in previous reviews (DeOreo & Keogh, 1980; Espenshade & Eckert, 1974; Wickstrom, 1983). We are hesitant to offer an explanation for this unexpected result because the 95% confidence intervals are

very large prior to puberty, indicating that effect sizes show large variations about the mean.

The curve for catching (Figure 5) is different from all the others. Boys' performance exceeds girls' performance in early childhood (3 years of age), but the differences are reduced to zero by 5 years of age. Boys' performance increasingly exceeds the girls' during elementary school until the difference is 1 standard deviation unit by age 13. These differences are likely environmental due to boys having more practice. The effect sizes are also likely to be underestimated because the ball is generally projected with little velocity in the tests reported.

In our opinion, any gender differences in performance prior to puberty on these 10 tasks (balance, catching, dash, grip strength, long jump, pursuit rotor tracking, shuttle run, sit-ups, tapping, and vertical jump) are mostly environmentally induced. The attitudes, expectations, and actions of parents, teachers, peers, and age-group coaches either produce or reinforce the differences. This is reflected by the increased opportunity and encouragement to practice these tasks afforded to boys. We believe that differences this small ($\overline{ES'} < 0.50$) could easily be eliminated if girls and boys were treated similarly.

Six tasks (dash, long jump, sit-ups, grip strength, shuttle run, and vertical jump) show postpuberty changes that are probably related to biological development, even though environmental factors continue to be important. Although equal encouragement and opportunity to practice would probably reduce gender differences after puberty, boys would still, on the average, perform better than girls. For balance, pursuit rotor tracking, and tapping, the

biological changes associated with puberty do not seem to be the cause of the slight increase observed. We suspect that environmentally related variables such as competitive motivation and sex-role expectations that become increasingly significant for boys and girls create these differences. The differences in catching skills after puberty are probably not a good estimate of the true difference in sports in which the ball is projected with velocity (e.g., baseball, softball).

Throwing is different. Even a casual glance at Figure 1 indicates that throwing for velocity and distance are different from the other 10 tasks. The differences are 1.5 standard deviation units at 3 years of age. These differences increase substantially during the elementary school years. We only have data points through 12 years of age for throwing velocity, but the boys' throwing velocity is already between 3.5 and 4 standard deviation units higher than the girls'. The acceleration of gender differences is not so rapid for distance thrown but is above 2 standard deviation units at age 12 and above 3 standard deviation units after age 16. The 95% confidence intervals are moderately tight on throwing velocity and very tight on throwing for distance. Thus, we have considerable confidence in these developmental curves.

Although Maccoby and Jacklin (1974) discussed some differences in the treatment of boys and girls in early childhood, differences as large as 1.5 standard deviation units at age 3 are unlikely to be completely environmentally caused. A substantial amount of throwing practice for boys would be needed to produce a difference this large. However, the fact that this relatively large gender difference continues to increase in childhood and adolescence is probably a combination of biology and environment.

Malina (1984) reported that prepubescent boys have slightly more total lean body mass and less fat in both an absolute and relative sense than girls. In addition, sex differences in somatotypes during early childhood have been reported (Walker, 1962). Over 50% of girls have a larger endomorphic component in body type (as compared with ectomorphic and mesomorphic), whereas over 50% of boys have a larger mesomorphic component. Malina (1975) believed these sex differences were of genetic origin. These total body differences are

also present specifically in the arm. Boys have a greater midarm circumference and a smaller triceps skinfold (i.e., more muscle tissue) than girls. Haubenstricker and Sapp (1980) reported that boys' forearms are 6 mm (on the average) longer than girls by 5.5 years, and that the differences continue throughout childhood. In addition, boys show an increasing advantage in late childhood in the biacromial/bicristal (shoulder/hip) ratio (Malina, 1984). Although each of these biological differences is small, taken together they may account for a portion of the gender difference in throwing performance. That biology is a factor in the differences is further confirmed by the fact that training in throwing has had little effect in reducing male-female performance differences in young children (Dusenberry, 1952; Halverson, Robertson, Safrit, & Roberts, 1977).

The increasing advantage boys gain over girls in throwing performance during elementary school is also partially attributable to environment. Both data (Halverson et al., 1982) and observation at playing fields indicate that elementary-age boys practice throwing skills much more than girls.

Thus, although gender differences in throwing velocity and throwing for distance can probably be reduced by providing equal encouragement and practice for girls, we believe that biological factors will not allow, on the average at least, their elimination. Therefore, prior to puberty, careful groupings by skill level should be made if boys and girls are to participate jointly in sports and skill drills in which throwing is important. This suggestion is reinforced by looking at the catching difference between boys and girls. The boys catch better during the elementary school years. The catching difference is most likely minimized by the low ball velocity reported in most tests of catching.

Of the 12 tasks that were age related, sports and skill drills involving throwing (or catching throws of high velocity) are the only ones for which biology appears to play an important role in the development of gender differences prior to puberty. Given the documented differences in expectations and treatment of boys and girls, the differences before puberty in the other 10 tasks are probably environmentally induced. After puberty, the importance of biology and environment are confounded in

many motor tasks, but especially in tasks requiring size, strength and endurance, and power.

Tasks Where Differences Were Not Age Related

Effect sizes for gender in 8 of the 20 (40%) motor performance tasks were not related to age. Thus, performance on these tasks did not change in any systematic way across childhood and adolescence. The gender differences favored boys and were low (0.01 to 0.38) for four of the tasks: agility, anticipation timing, arm hang, and reaction time. The differences were low and favored girls for fine eye-motor coordination (-0.21) and flexibility (-0.29). We see little reason to suspect that any of these small differences have a biological basis. The only difference of any substantial size—anticipation timing ($ES' = 0.38$)—is likely to be a result of boys, more often than girls, practicing sports (e.g., baseball, football) that require this skill.

The effect size of throwing for accuracy ($ES' = 0.96$), even though large, probably reflects more practice by boys rather than any biological differences. The difference in male and female performance is not developmental. The nature of the tests for throwing accuracy (i.e., the performer is a short distance from the target) is such that throwing velocity is not a factor.

The wall volley (batting an inflated ball repeatedly against a wall with the hands) also has a large effect size favoring boys. We suspect this reflects greater amounts of practice for the boys as well as their advantage in anticipation timing skills.

We see little reason to suspect that gender differences in performance on any of these eight tasks reflect anything but environmental factors. If biology were involved in the differences, we would expect them to be larger, especially after puberty.

Tasks in Which Characteristics Coded From Studies Were Important

In 3 of the 12 tasks in which gender differences were related to age, effect sizes were also related to other characteristics coded about the study. Both the shuttle run and sit-ups were negatively related to the gender of the first au-

thor. Thus, effect sizes were larger (favored boys) when the first author was female. If, in fact, the first author and the experimenter are the same person (not always true), the finding is consistent with Rikli's (1976) report that experimenters of the opposite gender cause larger gender differences in motor performance for tasks in which more effort results in better performance.

Effect sizes in the shuttle run and the dash were both positively related to whether the article was published or not, being larger in those published. This finding may reflect journals' bias toward publishing articles in which differences are significant.

Effect sizes in the shuttle run were also positively related to the validity of the study; more valid studies showed larger differences favoring boys. Sit-ups were negatively related to the year published, indicating that differences were larger in studies published before 1970. This may reflect the increasing standardization of sit-up tests. Earlier tests involved the number of sit-ups the performer could do; more recent tests typically have either a time limit (usually 30 s to 60 s) or a maximum number (50 to 100). Thus, differences between boys and girls would be reduced in more recent tests.

Issues That Remain Unaddressed

There are a number of important points that we have not considered here. First, motor performance is only the outcome of the movement. Although the outcome reflects the movement process, it does not do so perfectly and does not describe this process. Even a casual observer notes that individuals may run or throw in different ways, yet obtain similar outcomes or, conversely, run or throw in similar ways and obtain different outcomes. Thus, whether the development of gender differences across age exists in the quality of the movement cannot be determined by this meta-analysis, although work has been done in this area (see Robertson, 1984, for a summary). Robertson (1982) suggested that the form of the movement was not different between girls and boys, just that girls lag behind boys in the development of "good" form. This suggests that the nature of the underlying motor control mechanisms do not differ by gender. Although this view is consistent with most of the motor per-

formance tasks evaluated in this study (at least prior to puberty), it could be inconsistent with the large gender differences noted in throwing and in some of the other tasks after puberty.

A second point that remains unaddressed is the ethnicity and socioeconomic status of the subjects. We estimate that most subjects in the 64 studies included are white and of middle socioeconomic status. However, that information is not generally provided in the studies, particularly in the earlier ones. These issues may be of considerable importance because ethnic variation in rearing has been suggested as a factor in motor performance variation. Yet, ethnicity and socioeconomic status are confounded with gender in this report.

Third, the role of genetics (and its interaction with environment) cannot be evaluated from this data base. However, studies of twins and siblings suggest sex differences in the heritability of motor performance.

Last, an important issue is whether longitudinal and cross-sectional data points estimate gender differences in motor performance equally well at any given age. These data do not include enough longitudinal data points for any specific motor performance task to provide a fair test of this question. However, meta-analysis offers an interesting approach to this issue, if a task could be found in which a sufficient number of longitudinal and cross-sectional data points were available.

Conclusions

We believe the gender differences prior to puberty in 15 of the 20 tasks (agility, anticipation timing, arm hang, balance, dash, grip strength, fine eye-motor coordination, flexibility, long jump, pursuit rotor tracking, reaction time, shuttle run, sit-ups, tapping, vertical jump) studied are environmentally induced. This conclusion is based on the small effect sizes, usually less than 0.50 of a standard deviation unit, as well as documented observations that treatment, expectations, and practice opportunities differ by gender. If one single factor is of importance, it is that boys are involved in more competitive games than girls and generally participate in games of longer duration (Lever, 1976).

The effect sizes for 6 of the previously mentioned 15 tasks (dash, grip strength, long jump,

shuttle run, sit-ups, vertical jump) show rapid increases at puberty that are probably associated with the increase in boys' size and strength. However, the same environmental variables discussed previously inflate these gender differences beyond biological explanations.

The differences in effect sizes for throwing (velocity or distance) seem to begin with biological differences (1.5 standard deviation units at 3 years of age), but are increased by more practice opportunities for boys. The effect size for throwing accuracy is also large and may have some biological basis. However, given that the throwing-accuracy tests usually require minimal force, we suspect this gender difference in performance is also the result of boys practicing more than girls.

These findings of gender differences in motor performance are generally consistent with recent theories about sex-role development. Robinson and Green (1981) suggested that both cognitive-developmental (Kohlberg & Ullian, 1974) and transcendence (Hefner, Rebecca, & Oleshansky, 1975) views of sex-role development indicate that early development is strongly influenced by parents and peers. As children enter elementary school, teachers, peers, and parents influence sex stereotyping. Finally, as adults, men and women choose the type of sex role that fits their natures. However, biology does appear to play a greater role in gender differences for throwing at all ages and postpuberty in tasks for which size, strength, and power are important. This at least limits complete choice postpuberty in cross-sex, high-level competition for sports involving size, strength, and power. We are not suggesting that women cannot become skillful performers in these types of sports, just that some levels of joint sport participation remain constrained by the biology of gender differences.

References

- American Alliance for Health, Physical Education, Recreation and Dance. (1976). *Youth fitness test manual*. Washington, DC: Author.
- Anastasi, A. (1981). Sex differences: Historical perspectives and methodological implications. *Developmental Review*, 1, 187-206.
- Burmeister, W. (1965). Body cell mass as the basis of allometric growth functions. *Annales Paediatrici*, 204, 65-72.

- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Corbin, C. B. (Ed.). (1980). *A textbook of motor development* (2nd ed.). Dubuque, IA: Brown.
- DeOreo, K., & Keogh, J. (1980). Performance of fundamental motor tasks. In C. B. Corbin (Ed.), *A textbook of motor development* (2nd ed., pp. 76-91). Dubuque, IA: Brown.
- Dusenberry, L. M. (1952). A study of the effects of training in ball throwing by children ages three to seven. *Research Quarterly*, 23, 9-14.
- Espenschiede, A., & Eckert, H. (1974). Motor development. In W. R. Johnson & E. R. Buskirk (Eds.), *Science and medicine in exercise and sport* (2nd ed., pp. 322-333). New York: Harper & Row.
- Espenschiede, A. S., & Eckert, H. M. (1980). *Motor development* (2nd ed.). Columbus, OH: Merrill.
- Fagot, B. I. (1978). The influence of sex of child on parental reaction to toddler children. *Child Development*, 49, 459-465.
- Fling, S., & Manosevitz, M. (1972). Sex typing in nursery school children's play interest. *Developmental Psychology*, 7, 146-152.
- Frisancho, A. R. (1981). New norms of upper limb fat and muscle areas for assessment of nutritional status and weight. *American Journal of Clinical Nutrition*, 34, 2540-2545.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Greendorfer, S. L. (1980). Gender differences in physical activity. *Motor Skills: Theory into Practice*, 4, 83-90.
- Halverson, L. E., Robertson, M. A., & Langendorfer, S. (1982). Development of the overarm throw: Movement and ball velocity changes by seventh grade. *Research Quarterly for Exercise and Sport*, 53, 198-205.
- Halverson, L. E., Robertson, M. A., Safrit, M. J., & Roberts, T. W. (1977). Effect of guided practice on overhand-throw ball velocities of kindergarten children. *Research Quarterly*, 48, 311-318.
- Haubenstricker, J., & Sapp, M. (1980, April). *A longitudinal look at physical growth and motor performance: Implications for elementary and middle school activity programs*. Paper presented at the meeting of the American Alliance for Health, Physical Education, Recreation and Dance, Detroit, MI.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-128.
- Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics*, 7, 119-137.
- Hedges, L. V., & Olkin, I. (1983). Regression models in research synthesis. *American Statistician*, 37, 137-140.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Hefner, R., Rebecca, M., & Oleshansky, B. (1975). Development of sex-role transcendence. *Human Development*, 18, 143-158.
- Henry, F. M. (1968). Specificity vs. generality in learning motor skill. In R. C. Brown & G. S. Kenyon (Eds.), *Classical studies on physical activity* (pp. 328-331). Englewood Cliffs, NJ: Prentice-Hall.
- Housner, L. D. (1981). Sex-role stereotyping: Implications for teaching elementary physical education. *Motor Skills: Theory into Practice*, 5, 107-116.
- Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using ω^2 and d . *American Psychologist*, 36, 892-901.
- Jacklin, C. N. (1981). Methodological issues in the study of sex-related differences. *Developmental Review*, 1, 266-273.
- Johnson, B. L., & Nelson, J. K. (1979). *Practical measurement for evaluation in physical education* (3rd ed.). Minneapolis, MN: Burgess.
- Kohlberg, L., & Ullian, D. Z. (1974). Stages in the development of psychosexual concepts and attitudes. In R. C. Friedman, R. M. Richart, & R. L. Vande Wiele (Eds.), *Sex differences in behavior* (pp. 209-222). New York: Wiley.
- Lansky, L. M. (1967). The family structure also affects the model: Sex-role attitudes in parents of preschool children. *Merrill-Palmer Quarterly*, 13, 139-150.
- Lever, J. (1976). Sex differences in the games children play. *Social Problems*, 23, 478-487.
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Malina, R. M. (1975). *Growth and development: The first twenty years in man*. Minneapolis, MN: Burgess.
- Malina, R. M. (1984). Physical growth and maturation. In J. R. Thomas (Ed.), *Motor development during childhood and adolescence* (pp. 2-26). Minneapolis, MN: Burgess.
- Malina, R. M., & Johnson, F. E. (1967). Significance of age, sex, and maturity differences in upper arm composition. *Research Quarterly*, 38, 219-230.
- National Center for Health Statistics. (1977). NCHS growth curves for children, birth-18 years, United States. *Vital and Health Statistics*, Series 11, No. 165.
- Ridenour, M. V. (Ed.). (1978). *Motor development: Issues and applications*. Princeton, NJ: Princeton Book.
- Rikli, R. (1976). Physical performance scores as a function of experimenter sex and experimenter bias. *Research Quarterly*, 47, 776-782.
- Robertson, M. A. (1982). Describing "stages" within and across motor tasks. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination* (pp. 293-307). New York: Wiley.
- Robertson, M. A. (1984). Changing motor patterns during childhood. In J. R. Thomas (Ed.), *Motor development during childhood and adolescence* (pp. 48-90). Minneapolis, MN: Burgess.
- Robinson, B. E., & Green, M. G. (1981). Beyond androgyny: The emergence of sex-role transcendence as a theoretical construct. *Developmental Review*, 1, 247-265.
- Roche, A. F., & Malina, R. M. (Eds.). (1983). *Manual of physical status and performance in childhood* (Vol. 1). New York: Plenum.
- Schmidt, R. A. (1982). *Motor control and learning*. Champaign, IL: Human Kinetics.
- Sherif, C. W., & Rattray, G. D. (1976). Psychological development and activity in middle childhood. In J. G. Albinson & G. M. Andrew (Eds.), *Child in sport and physical activity* (pp. 97-132). Baltimore: University Park.
- Thomas, J. R. (Ed.). (1984). *Motor development during childhood and adolescence*. Minneapolis, MN: Burgess.

- Thomas, J. R., & Nelson, J. K. (1985). *Introduction to research in HPERD*. Champaign, IL: Human Kinetics.
- Tolson, H. (1980). An adjunct to statistical significance: ω^2 . *Research Quarterly for Exercise and Sport*, 51, 580-584.
- Walker, R. M. (1962). Body build and behavior in young children: I. Body build and nursery school teachers' ratings. *Monographs of the Society for Research in Child Development*, 27 (3, Serial No. 34).
- Wickstrom, R. L. (1983). *Fundamental motor patterns* (3rd ed.). Philadelphia: Lea & Febiger.

Appendix

Studies Included in the Meta-Analysis ($N = 64$)

- Bachman, J. C. (1961). Motor learning and performance as related to age and sex in two measures of balance coordination. *Research Quarterly*, 32, 123-137.
- Boley, E. H. (1975). *Generality and specificity of motor performance of children as related to age and sex*. Unpublished master's thesis, Louisiana State University, Baton Rouge, LA.
- Broekhoff, J. (1978). Longitudinal comparison of the growth, physical fitness, and motor performance of suburban and inner city elementary school children. In F. Landry & W. A. R. Orban (Eds.) *Motor learning, sport psychology, pedagogy, and didactics of physical activity* (pp. 203-210). Miami, FL: Symposia Specialists.
- Caskey, S. R. (1968). Effects of motivation on standing broad jump performance of children. *Research Quarterly*, 39, 54-59.
- Cowgill, S. (1978). *Nebraska physical fitness norms: Grades 1-12*. Unpublished master's thesis, Colorado State University, Fort Collins, CO.
- Davol, S. H., & Breakell, S. L. (1968). Sex differences in rotary pursuit performance of young children: A follow-up. *Perceptual and Motor Skills*, 26, 1199-1202.
- Davol, S. H., Hastings, M. L., & Klein, D. A. (1965). Effect of age, sex, and speed of rotation on rotary pursuit performance by young children. *Perceptual and Motor Skills*, 21, 351-357.
- Dohrman, P. (1964). Throwing and kicking ability of 8-year-old boys and girls. *Research Quarterly*, 35, 464-471.
- Dunham, P., Jr. (1977). Age, sex, speed, and practice in coincidence-anticipation performance of children. *Perceptual and Motor Skills*, 45, 187-193.
- Eckert, H. M. (1970). Visual-motor tasks at 3 and 4 years of age. *Perceptual and Motor Skills*, 31, 560.
- Eckert, H. M. (1974). Variability in skill acquisition. *Child Development*, 45, 487-489.
- Eckert, H. M., & Eichorn, D. H. (1974). Construct standard in skilled action. *Child Development*, 45, 439-445.
- Eckert, H. M., & Eichorn, D. H. (1977). Developmental variability in reaction time. *Child Development*, 48, 452-458.
- Eckert, H. M., & Rarick, G. L. (1976). Stabilometer performance of educable mentally retarded and normal children. *Research Quarterly*, 47, 619-623.
- Espenschade, A. S., & Meleney, H. E. (1961). Motor performance of adolescent boys and girls of today in comparison with those of 24 years ago. *Research Quarterly*, 32, 186-189.
- Finlayson, M. A. J., & Reitan, R. M. (1976). Handedness in relation to measures of motor and tactile-perceptual functions in normal children. *Perceptual and Motor Skills*, 43, 475-481.
- Fitch, J. H. (1980). *An analysis of the factors on the North Carolina motor fitness battery test among students aged ten to seventeen*. Unpublished master's thesis, North Carolina Central, Durham, NC.
- Gabbard, C., Kirby, T., & Patterson, R. (1979). Reliability of the straight-arm hang for testing muscular endurance among children 2 to 5. *Research Quarterly*, 50, 735-738.
- Glassow, R. B., Halverson, L. E., & Rarick, G. L. (1965). *Improvement of motor development and physical fitness in elementary school children*. Unpublished manuscript, University of Wisconsin, Cooperative Research Project #696.
- Glover, E. G. (1962). *Physical fitness test items for boys and girls in the first, second, and third grades*. Unpublished master's thesis, University of North Carolina, Chapel Hill, NC.
- Govatos, L. A. (1959). Relationships and age differences in growth measures and motor skills. *Child Development*, 30, 333-340.
- Green, J. M. (1973). *The relative effectiveness of a perceptual motor program, a movement education program, and a traditional program in the enhancement of motor performance of kindergarten children*. Unpublished master's thesis, University of Washington, Seattle, WA.
- Greene, J. L. (1973). *Effects of a prescribed physical education program upon movement characteristics of 4-year-old boys and girls*. Unpublished

- doctoral dissertation, University of Utah, Salt Lake City, UT.
- Halverson, L. E., Robertson, M. A., & Langendorfer, S. (1982). Development of the overarm throw: Movement and ball velocity changes by seventh grade. *Research Quarterly*, 53, 198-205.
- Halverson, L. E., Robertson, M. A., Safrit, M. J., & Roberts, T. W. (1977). Effect of guided practice on overhand-throw ball velocities of kindergarten children. *Research Quarterly*, 48, 311-318.
- Harper, C. J. (1975). *Movement responses of kindergarten children to a change of direction task—an analysis of selected measures*. Unpublished master's thesis, University of Wisconsin, Madison, WI.
- Hartman, D. (1943). The hurdle jump as a measure of the motor proficiency of young children. *Child Development*, 14, 201-211.
- Harvey, D. A. (1970). *The effects of level of aspiration and team competition as motivational techniques upon children's performances on selected sports skill tests*. Unpublished doctoral dissertation, Indiana University, Bloomington, IN.
- Haywood, K. M., Greenwald, G., & Lewis, C. (1981). Contextual factors and age group differences in coincidence-anticipation performance. *Research Quarterly*, 52, 458-464.
- Hunsicker, P. A. (1965). *A survey and comparison of youth fitness, 1958-1965*. Unpublished manuscript, University of Michigan, Cooperative Research Project #2418.
- Ikedo, N. (1961). *A comparison of physical fitness of children in Iowa, USA and Tokyo, Japan*. Unpublished doctoral dissertation, State University of Iowa, Iowa City, IA.
- Ingersoll, M. T. (1976). *The motor fitness of primary boys and girls*. Unpublished master's thesis, Ithaca College, Ithaca, NY.
- Isaacs, L. D. (1980). Effects of ball size, ball color, and preferred color on catching by young children. *Perceptual and Motor Skills*, 51, 583-586.
- Kane, R. J., & Meredith, H. V. (1952). Ability in the standing broad jump of elementary school children 7, 9, and 11 years of age. *Research Quarterly*, 23, 198-208.
- Keating, D. P., & Bobbitt, B. L. (1978). Individual and developmental differences in cognitive-processing components of mental ability. *Child Development*, 49, 155-167.
- Keogh, J. (1965). *Motor performance of elementary school children*. Unpublished manuscript, Department of Physical Education, University of California, Los Angeles. (Available through ERIC)
- Knights, R. M., & Moulds, A. D. (1967). Normative and reliability data on finger and foot tapping in children. *Perceptual and Motor Skills*, 25, 717-720.
- Latchaw, M. (1954). Measuring selected motor skills in fourth, fifth, and sixth grades. *Research Quarterly*, 25, 439-449.
- Lee, A. M., Fant, H., Life, M. L., Lipe, L., & Carter, J. A. (1978). Field independence and performance on ball handling tasks. *Perceptual and Motor Skills*, 46, 439-442.
- Maples, M. G. (1977). *Second grade children's performance on the overhand throw in relation to maternal and self preference for play activities*. Unpublished master's thesis, Purdue University, West Lafayette, IN.
- McCaskill, C. L., & Wellman, B. L. (1938). A study of common motor achievements of the preschool ages. *Child Development*, 9, 141-149.
- Miller, J. L. (1957). Effect of instruction on development of throwing accuracy of first grade children. *Research Quarterly*, 28, 132-137.
- Milne, C., Seefeldt, V., & Reuschlein, P. (1976). Relationship between grade, sex, race, and motor performance in young children. *Research Quarterly*, 47, 726-730.
- Miyashita, M., & Kanehisa, H. (1979). Dynamic peak torque related to age, sex, and performance. *Research Quarterly*, 50, 249-255.
- Montpetit, R. R., Montoye, H. J., & Laeding, L. (1967). Grip strength of school children, Saginaw, Michigan: 1899 and 1964. *Research Quarterly*, 38, 231-240.
- Morris, A. M., Williams, J. L., Atwater, A. E., & Wilmore, J. H. (1982). Age and sex differences in motor performance of 3 through 6-year-old children. *Research Quarterly*, 53, 214-221.
- Nestroy, J. A. (1978). *Fitness levels of children taught by the physical education specialists and classroom teachers*. Unpublished master's thesis, Texas Woman's University, Denton, TX.
- Pissanos, B. W., Moore, J. B., & Reeve, T. G. (1983). Age, sex, and body composition as predictors of children's performance on basic motor abilities and health-related fitness items. *Perceptual and Motor Skills*, 56, 71-77.
- Pomeroy, J. E. (1938). The relation of reaction time of five-year-old children to various factors. *Child Development*, 9, 281-283.
- Robertson, M. A., Halverson, L. E., Langendorfer, S., & Williams, K. (1979). Longitudinal changes in children's overarm throw ball velocities. *Research Quarterly*, 50, 256-264.
- Ross, B. M. (1960). A study of the performance of boys and girls taught by the specialist and nonspecialist. *Research Quarterly*, 31, 199-207.
- Seils, L. G. (1951). The relationship between measures of physical growth and gross motor performance of primary-grade school children. *Research Quarterly*, 22, 244-258.

Singer, R. N. (1969). Physical characteristic, perceptual-motor, and intelligence differences between third- and sixth-grade children. *Research Quarterly*, 40, 803-811.

Smith, J. (1956). Relation of certain physical traits and ability of motor learning in elementary children. *Research Quarterly*, 27, 221-228.

Smith, T. L. (1982). Self-concepts and movement skills of third grade children after physical education programs. *Perceptual and Motor skills*, 54, 1145-1146.

Smoll, F. L. (1966). *The influence of physical growth and muscular strength upon motor performances within and between year observations*. Unpublished master's thesis, University of Wisconsin, Madison, WI.

Stachnik, T. J. (1964). Cross-validation of psychomotor test for children. *Perceptual and Motor Skills*, 18, 913-916.

Taddonio, D. A. (1966). Effect of daily fifteen-minute periods of calisthenics upon the physical fitness of fifth grade boys and girls. *Research Quarterly*, 37, 276-281.

Thomas, J. R., Gallagher, J. D., & Purvis, G. J. (1981). Reaction time and anticipation time: Effects of development. *Research Quarterly*, 52, 359-367.

Torres, J. A. (1966). *The relationship between figure-ground perceptual ability and ball catching ability in ten and thirteen year old boys and girls*. Unpublished master's thesis, Purdue University, West Lafayette, IN.

Trussell, E. M. (1969). Relation of performance of selected physical skills to perceptual aspects of reading readiness in elementary school children. *Research Quarterly*, 40, 383-390.

Wilson, J. G., Silva, P. A., & Williams, S. M. (1981). An assessment of motor ability in seven year olds. *Journal of Human Movement Studies*, 7, 221-231.

Workman, D. J. (1979). Comparison of performance of children taught by the physical education specialist and by the classroom teacher. *Research Quarterly*, 49, 389-394.

Wright, E. J. (1967). Effects of light and heavy equipment on acquisition of sports-type skills by young children. *Research Quarterly*, 38, 705-714.

Received October 17, 1984
Revision received February 7, 1985 ■

Editorial Consultants for This Issue: Review Articles

Henry E. Adams	Robert W. Frick	Lovick C. Miller, Jr.
Stephen M. Auerbach	Robert J. Gatchel	Gerald M. Murch
V. Edwin Bixenstein	Claire Golomb	Michael T. Nietzel
Robert Boice	Susan M. Gore	William H. Norman
Arthur P. Brief	J. Richard Hackman	Joy D. Osofsky
John C. Brigham	John F. Hall	Anne Petersen
Bert R. Brown	Ronald K. Hambleton	Kvetoslav Prazdny
Marcelline M. Burns	Anne S. Hardesty	Lynn P. Rehm
John A. Carpenter	D. Dwight Harshbarger	Richard R. Reilly
Dianne L. Chambless	Richard M. Held	James A. Russell
George A. Clum	Jonathan S. Kellerman	Dennis C. Russo
William C. Coe	Gary Kose	Lawrence J. Siegel
John D. Cone	Lawrence A. Kurdek	Marvin Siegelman
Lynn A. Cooper	Benjamin B. Lahey	Mary V. Solanto
Paul T. Costa, Jr.	Janet T. Landman	Philip Spergel
James C. Coyne	Harold Leitenberg	Helen B. Tager-Flusberg
Joseph W. Critelli	Edward L. Levine	Hendrika Vande Kemp
Kenneth A. Dodge	Laura C. Leviton	James W. Varni
Virginia I. Douglas	Robert M. Malina	Michael Weissberg
Barbara A. Everett	Frank T. Masur, III	Gary L. Wells
Stanley Feldstein	John P. Meyer	Donald A. Williamson
David Foulkes		

RESEARCH ARTICLE

Performance Development in Adolescent Track and Field Athletes According to Age, Sex and Sport Discipline

Espen Tønnessen^{1*}, Ida Siobhan Svendsen², Inge Christoffer Olsen³, Atle Guttormsen⁴, Thomas Haugen¹

1 Norwegian Olympic Federation, Oslo, Norway, **2** School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom, **3** Diakonhjemmet, Oslo, Norway, **4** University of Life Sciences, Ås, Norway

* espen.tonnessen@olympiatoppen.no



OPEN ACCESS

Citation: Tønnessen E, Svendsen IS, Olsen IC, Guttormsen A, Haugen T (2015) Performance Development in Adolescent Track and Field Athletes According to Age, Sex and Sport Discipline. PLoS ONE 10(6): e0129014. doi:10.1371/journal.pone.0129014

Academic Editor: Nir Eynon, Victoria University, AUSTRALIA

Received: February 12, 2015

Accepted: May 3, 2015

Published: June 4, 2015

Copyright: © 2015 Tønnessen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are publically available online: <http://www.friidrett.no/stevner/statistikk1/alle%20tidere/Sider/Norgesbestegjennomalletider.aspx>

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Introduction

Sex-specific differences that arise during puberty have a pronounced effect on the training process. However, the consequences this should have for goal-setting, planning and implementation of training for boys and girls of different ages remains poorly understood. The aim of this study was to quantify performance developments in athletic running and jumping disciplines in the age range 11-18 and identify progression differences as a function of age, discipline and sex.

Methods

The 100 all-time best Norwegian male and female 60-m, 800-m, long jump and high jump athletes in each age category from 11 to 18 years were analysed using mixed models with random intercept according to athlete.

Results

Male and female athletes perform almost equally in running and jumping events up to the age of 12. Beyond this age, males outperform females. Relative annual performance development in females gradually decreases throughout the analyzed age period. In males, annual relative performance development accelerates up to the age of 13 (for running events) or 14 (for jumping events) and then gradually declines when approaching 18 years of age. The relative improvement from age 11 to 18 was twice as high in jumping events compared to running events. For all of the analyzed disciplines, overall improvement rates were >50% higher for males than for females. The performance sex difference evolves from < 5% to 10-18% in all the analyzed disciplines from age 11 to 18 yr.

Conclusion

To the authors' knowledge, this is the first study to present absolute and relative annual performance developments in running and jumping events for competitive athletes from early

to late adolescence. These results allow coaches and athletes to set realistic goals and prescribe conditioning programs that take into account sex-specific differences in the rate of performance development at different stages of maturation.

Introduction

Running and jumping are fundamental motor skills in most sport disciplines. Such skills are developed throughout life via growth, maturation and training [1]. Males have consistently performed better in sports reliant on sprinting, jumping and endurance capacity [2–6]. As women have gradually been given the opportunity to compete on equal terms with men, the sex difference in performance has progressively narrowed before stabilizing in the 1980s [2,4,6]. Recent studies of world-class athletes indicate that the sex difference is 10–12% for running events [2,4,6] and ~19% for jumping events [6]. To date, no studies have documented when these sex differences arise. An understanding of the expected sex differences is important to allow coaches and athletes to set realistic goals and prescribe optimal conditioning programs that take into account sex-specific differences in the rate of performance development at different stages of maturation.

Sexual dimorphism during puberty is highly relevant for understanding sex-specific performance developments in sports. The initiation of the growth spurt in well-nourished girls occurs at about 9–10 yrs of age. Age at peak height velocity (PHV) and peak weight velocity (PWV) in girls is 11–12 and 12–13 yrs, respectively, with an average 7–9 cm and 6–9 kg annual increase [7,8,9]. The growth spurt and PHV in girls occurs approximately 2 years earlier than for boys. However, the magnitude of the growth spurt is typically greater in boys, as they on average gain 8–10 cm and 9–10 kg annually at PHV and PWV, respectively. Girls experience an escalation in fat mass compared to boys. Fat free mass (FFM) (also termed lean muscle mass) is nearly identical in males and females up to the age of 12–13 yrs. FFM plateaus in females at 15–16 years of age, but continues increasing in males up to the age of 19–20 yrs [7,8,9]. On average, boys and girls increase their FFM by 7.2 and 3.5 kg·year⁻¹, respectively, during the interval near peak height velocity. Corresponding estimates for changes in absolute fat mass are 0.7 and 1.4 kg·year⁻¹, while estimates for relative fatness are -0.5% and +0.9%·year⁻¹ in boys and girls, respectively [8].

Previous cross-sectional studies of primarily non-competitive adolescents reveal that sex differences in physical capacities (assessed as $\dot{V}O_2$ peak or isometric strength in the majority of cases) are negligible prior to the onset of puberty [9–15]. During the adolescent growth spurt, however, marked sex differences develop. This can primarily be explained by hormone-dependent changes in body composition [8,16] and increased red blood cell mass in boys [16,17]. Despite such sex-specific differences having a pronounced effect on the training process, the consequences these dissimilarities should have for goal setting, planning and implementation of training for boys and girls of different ages, remains poorly understood. No published studies to date have investigated expected annual performance developments in running and jumping disciplines among competitive athletes, from the start through to the end of puberty.

Information regarding the realistic potential for development for a dedicated athlete from the start to the end of puberty is currently lacking in the research literature. Only longitudinal or mixed longitudinal data can provide adequate information regarding performance development during the adolescence growth spurt [14]. The Statistics Committee of the Norwegian Athletics Federation has, over several decades, recorded and systemized competition results from athletic events for all Norwegian athletes from the age of 11 years. This unique database

provides the opportunity to investigate the long-term performance development of competitive athletes in sports which place high demands on endurance running, sprint running and jumping ability, from childhood up until the end of puberty. Hence, the purpose of this study was to quantify performance developments in athletic running and jumping disciplines in the age range 11–18 yr and identify possible progression differences as a function of age, discipline and sex.

Materials and Methods

Data sample

This study was conducted in accordance with the declaration of Helsinki. Since our data are based on publically available resources, no informed consent was obtained. This study was approved by the local ethics committee at the Norwegian University of Life Sciences.

The Norwegian Athletics Association annually publishes all-time best results categorized by sex, age and discipline [18]. Each record within these rankings documents performance, name, birth date, club, competition date and venue where the result was set. The database is restricted to the individual season best for each discipline. In the present study, the 100 all-time best male and female 60 m, 800 m, long jump and high jump athletes in each age category from 11 to 18 years were included for analysis. According to Norwegian track & field regulations, athletes are allowed to compete in these disciplines from the age of 11. Long distance disciplines were excluded from analyses as Norwegian athletes are not allowed to compete in such disciplines before the age of 15. Throwing and hurdling events were also excluded due to different competition regulation standards across age categories (i.e. weight of throwing implement, hurdle height, etc.). The included 100 all-time best lists contain data back to 1975. Only outdoor season data were included for 800 m, long jump and high jump, while 60 m included both indoor and outdoor data. However, 60 m results obtained without electronic timing were excluded. Only 60 m and long jump results obtained with legal wind speed ($\leq 2 \text{ m}\cdot\text{s}^{-1}$) were included. Overall, the sample consisted of results from 1373 male and 1149 female athletes and a total of 6400 individual results. The distribution of male/female athletes across the analyzed sports disciplines were 408/369 for 60 m, 414/379 for 800 m, 440/371 for long jump and 354/347 for high jump. About 15–20% of the athletes were in the 100 all-time best lists in more than one discipline.

Statistical analyses

The data were analyzed using mixed models with result as dependent variable and age as fixed explanatory variable with an athlete-specific random variable to account for within-athlete dependency. Separate analyses were performed by discipline and sex. We used estimated marginal means with 95% confidence limits (95% CIs) to produce plots of expected progression by gender and exercise. Descriptive statistics are presented as mean and standard deviation (SD) in addition to percent (%) for change. Effect magnitudes (based on Cohen's *d*) across categories were interpreted categorically as small (*d* from 0.2 to 0.6), moderate (*d* from 0.6 to 1.2) or large (*d* from 1.2 to 2.0) using the scale presented by Hopkins et al. [19]. All analyses were performed using SAS software version 9.2 (SAS Institute Inc., Cary, NC, USA).

Results

60 m sprint

[Table 1](#) and [Fig 1](#) show that boys improve 0.3–0.5 s over 60 m sprint each year up to the age of 14 yr (very large to nearly perfect annual effect), 0.1–0.2 s annually from 14 to 17 yr (moderate

Table 1. Expected progressions in running and jumping performance for 11–18 yr old males and females.

Age (yr)	60 m		800 m		Long Jump		High Jump	
	Boys Progression (s and %)	Girls Progression (s and %)	Boys Progression (s and %)	Girls Progression (s and %)	Boys Progression m (%)	Girls Progression m (%)	Boys Progression m (%)	Girls Progression m (%)
11–12	-0.35 (4.1)	-0.35 (4.0)	-6.4 (4.4)	-7.3 (4.8)	+0.35 (7.4)	+0.36 (7.9)	+0.11 (7.4)	+0.10 (7.2)
12–13	-0.48 (5.8)	-0.25 (2.9)	-8.7 (6.2)	-5.5 (3.8)	+0.43 (8.6)	+0.30 (6.0)	+0.12 (7.9)	+0.09 (6.3)
13–14	-0.29 (3.7)	-0.16 (2.0)	-5.9 (4.5)	-3.6 (2.6)	+0.50 (9.0)	+0.21 (4.1)	+0.13 (8.1)	+0.06 (3.6)
14–15	-0.10 (1.3)	-0.02 (0.2)	-5.2 (4.1)	-2.2 (1.6)	+0.34 (5.6)	+0.13 (2.4)	+0.08 (4.3)	+0.04 (2.4)
15–16	-0.17 (2.3)	-0.08 (1.0)	-3.2 (2.7)	-1.6 (1.2)	+0.28 (4.4)	+0.10 (1.8)	+0.07 (3.6)	+0.03 (1.8)
16–17	-0.10 (1.4)	-0.07 (0.8)	-2.3 (1.9)	-1.5 (1.2)	+0.19 (2.9)	+0.06 (1.1)	+0.05 (2.5)	+0.01 (0.6)
17–18	-0.05 (0.7)	-0.02 (0.2)	-1.5 (1.4)	-0.6 (0.4)	+0.17 (2.5)	+0.02 (0.4)	+0.04 (1.9)	+0.01 (0.5)

Data are mean (standard deviation) for top 100 Norwegian male and female performers in each discipline.

doi:10.1371/journal.pone.0129014.t001

to large annual effect), and 0.05 s from age 17 to 18 yr (moderate effect). Relative annual improvement peaks between 12 and 13 yr (5.8%; nearly perfect effect), and then gradually declines to 0.7% between age 17 and 18 yr (moderate effect). On average, boys improve their 60 m performance by 18% from age 11 to 18 yr (Fig 2). Girls improve 0.35 s over 60 m from age 11 to 12 yr (4%; very large effect) (Table 1 and Fig 1). Then, absolute and relative annual improvement gradually slows and almost plateaus between age 14 and 15 (0.02 s; 0.2%; trivial effect). From age 15 to 17, annual improvement increases somewhat to 0.07–0.08 s (~1%; moderate effect) before plateauing again between age 17 and 18 (0.02 s; 0.2%; trivial effect). In total, girls improve their 60-m performance by 11% from age 11 to 18 yr (Fig 2). Fig 3 (panel A)

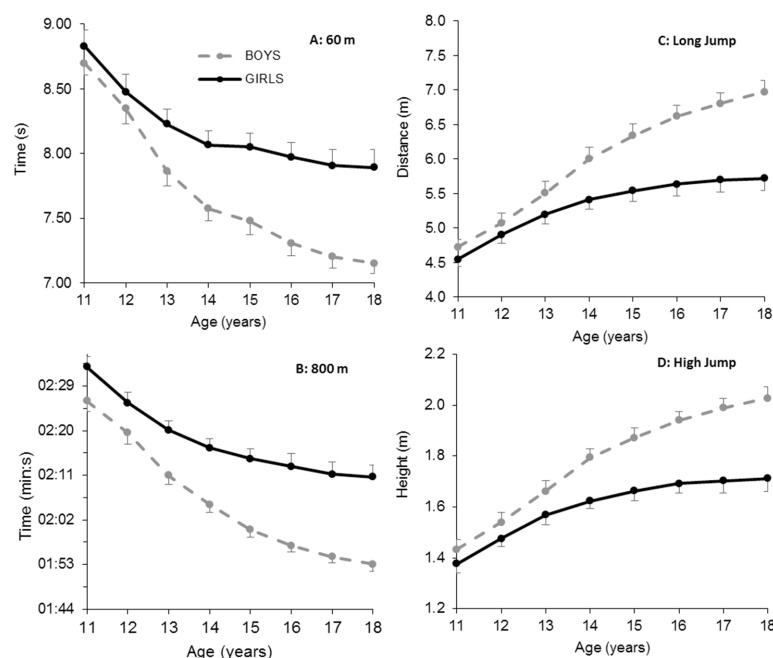


Fig 1. Performance development from age 11 to 18 in running and jumping disciplines. Data are mean $\pm$ SD for 60 m (panel A), 800 m (panel B) long jump (panel C) and high jump (panel D) for top 100 Norwegian male and female performers in each discipline.

doi:10.1371/journal.pone.0129014.g001

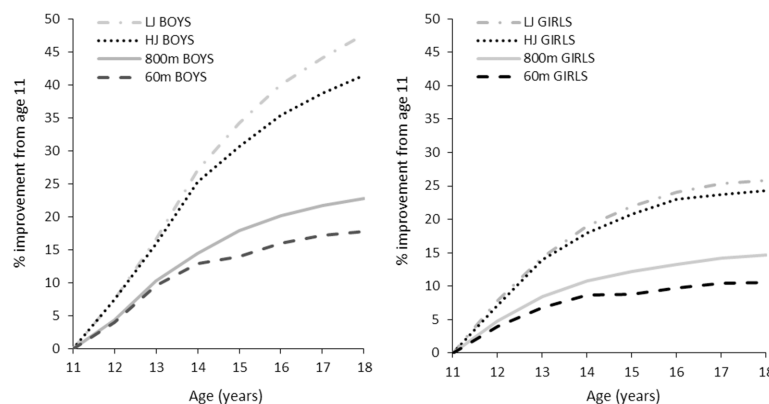


Fig 2. Percentage improvement in performance from age 11 to 18 in long jump, high jump, 60 m sprint and 800 m. Data are mean for top 100 Norwegian male (panel A) and female (panel B) performers in each discipline.

doi:10.1371/journal.pone.0129014.g002

shows that the sex difference for 60 m sprint evolves from 1.5% at age 11 to 10.3% at the age of 18. Table 2 presents sex ratio in performance at each age stage, and shows that the sex ratio for 60 m running performance develops from 0.99 at age 11 to 0.91 at age 18.

800 m

Table 1 and Fig 1 show that boys improve 6–9 s over 800 m each year up to age 14 yr (very large to nearly perfect annual effect). Relative annual improvement peaks between age 12 and 13 (6.2%; nearly perfect effect), then gradually decreases to 1.5 s between age 17 and 18 (1.4%; moderate effect). On average, boys enhance their 800-m performance by 23% from age 11 to 18 (Fig 2). For girls, both absolute and relative annual performance development gradually decreases across the analysed age stages. The improvement is slightly above 7 s between age 11 and 12 yr (4.8%; very large effect), decreasing to only 0.6 s from age 17 to 18 (0.4%; small effect) (Table 1 and Fig 1). Fig 2 shows that girls enhance their 800-m performance by 15% from age 11 to 18. The 800 m performance sex difference evolves from 4.8% at the age of 11 to 15.7% at the age of 18 (Fig 3, panel B). Table 2 shows that the sex ratio for 800 m running performance develops from 0.95 at age 11 to 0.86 at age 18.

Long jump

Table 1 and Fig 1 show that annual long jump improvement among boys gradually increases from 35 cm between age 11 and 12 yr (7.4%; very large effect) to 50 cm between age 13 and 14 (9%; very large effect). Both absolute and relative annual development then gradually falls to 17 cm between age 17 and 18 (2.5%; moderate effect). Fig 2 shows that boys, on average, improve their long jump performance by 48% from age 11 to 18 yr. For girls, both absolute and relative annual performance enhancement gradually falls from age 11 to 12 yr (36 cm; 7.9%; very large effect) until nearly plateauing between 17 and 18 yr (2 cm; 0.4%; trivial effect) (Table 1 and Fig 1). Overall, girls typically improve their long jump performance by 26% throughout the analysed age stages (Fig 2). The sex difference in long jump evolves from 3.6% at the age of 11 to 18% at the age of 18 (Fig 3, panel C). Table 2 shows that the sex ratio for long jump performance develops from 0.96 at age 11 to 0.82 at age 18.

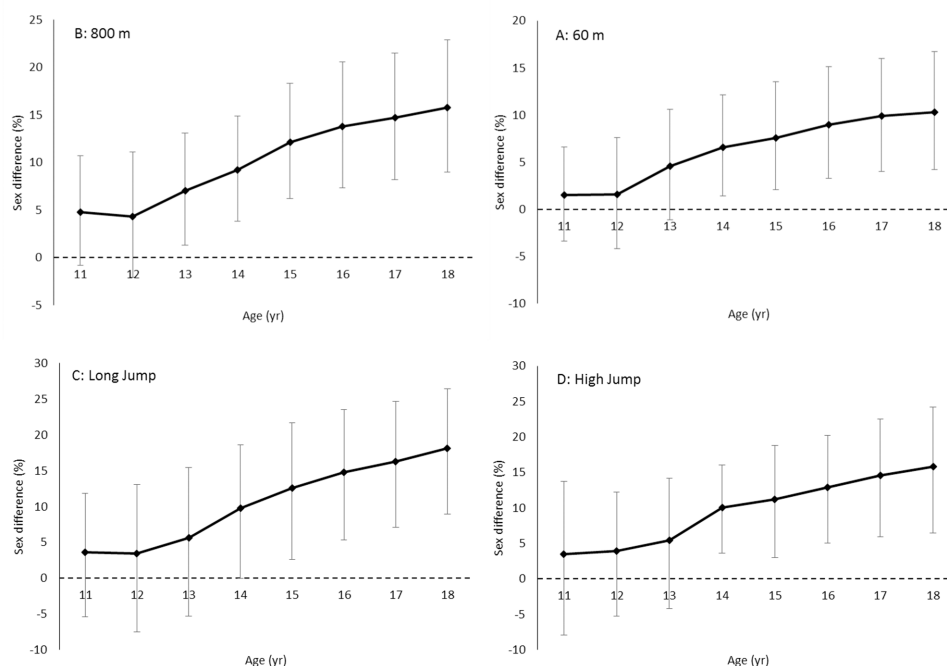


Fig 3. Sex difference (%) for performance in running and jumping disciplines from age 11 to 18. Data are mean and 95% CIs for 60 m (panel A), 800 m (panel B), long jump (panel C) and high jump (panel D) for top 100 Norwegian male and female performers in each discipline.

doi:10.1371/journal.pone.0129014.g003

High jump

Table 1 and Fig 1 show that boys improve their high jump performance by 11–13 cm each year up to the age of 14 (7–8%; very large annual effects). Both absolute and relative annual improvement peaks between age 13 and 14 (13 cm; 8.1%; very large effect), then gradually decreases to 4 cm from age 17 to 18 (1.9%; moderate annual effect). Overall, boys improve their high jump performance by, on average, 41% from age 11 to 18 (Fig 2). For girls, both absolute and relative annual improvement decreases from 10 cm from age 11 to 12 yr (7.2%; very large effect) until it plateaus from age 16 (1 cm; ~0.5%; small annual effects) (Table 1 and Fig 1). Overall, girls typically improve their high jump performance by 24% from age 11 to 18 (Fig 2).

Table 2. Sex ratio in running and jumping performance for 11–18 yr old males and females.

	60 m	800 m	Long Jump	High Jump
11	0.99	0.95	0.96	0.97
12	0.98	0.96	0.97	0.96
13	0.96	0.93	0.94	0.95
14	0.94	0.92	0.90	0.90
15	0.93	0.89	0.87	0.89
16	0.92	0.88	0.85	0.87
17	0.91	0.87	0.84	0.85
18	0.91	0.86	0.82	0.84

Data are calculated from mean results of top 100 Norwegian male and female performers in each discipline.

doi:10.1371/journal.pone.0129014.t002

The sex difference in high jump performance evolves from 3.5% at the age of 11 to 16% at the age of 18 (Fig 3, panel D). Table 2 shows that the sex ratio for high jump performance develops from 0.97 at age 11 to 0.84 at age 18.

Discussion

To the authors' knowledge, this is the first study to present absolute and relative annual performance developments in running and jumping events for competitive athletes from early to late adolescence. Data from the 100 best performers in each age category from age 11 to 18 years show that male and female athletes perform almost equally up to the age of 12. Beyond this age, males outperform females. Relative annual performance development in females gradually decreases throughout the analyzed age period. In males, annual relative performance development accelerates up to the age of 13 (for running events) or 14 (for jumping events) and then gradually declines when approaching 18 years of age. The relative improvement from age 11 to 18 was twice as high in jumping events compared to running events. For all of the analyzed disciplines, overall improvement rates were >50% higher for males than for females.

Sexual dimorphism during puberty is central in understanding the sex- and discipline-specific performance developments observed in the current study. During puberty, boys begin to produce higher levels of circulating testosterone. This affects the production of muscle fibers through direct stimulation of protein synthesis [16,17]. Higher testosterone levels result in more muscle mass, which in turn facilitates greater power production and more advantageous ground reaction forces during running and jumping. Adolescent weight gain in boys is principally due to increased height (skeletal tissue) and muscle mass, while fat mass remains relatively stable [7,8,9]. In contrast, during puberty girls begin to produce higher levels of circulating estrogen and other female sex hormones [16]. Compared to their male counterparts, they experience a less pronounced growth spurt and a smaller increase in muscle mass, but a continuous increase in fat mass, thereby lowering the critical ratio between muscular power and total body mass. Fat free mass (FFM) in males and females is nearly identical up to 12–13 years of age. Subsequently, FFM plateaus in females at 15–16 years of age, but continues to increase in males up to 19–20 years of age [7,8,9]. On average, boys and girls increase their FFM by 7.2 and 3.5 kg·year⁻¹, respectively, during the interval near peak height velocity. Corresponding estimates for changes in absolute fat mass are 0.7 and 1.4 kg·year⁻¹, while estimates for body fat percentage are -0.5% and +0.9%·year⁻¹ in boys and girls, respectively [8]. Moreover, increased red blood cell mass in boys at this time may contribute towards the widening sex difference for 800 m running [16,17]. Taking present and previous findings together, there appears to be a strong mechanistic connection between the observed sex-specific performance developments and hormone-dependent changes in body composition during puberty. Figs 1 and 2 are very consistent with the muscle mass proportion curve outlined by Malina et al. [8], where a sex difference break point at the age of 12 is clearly present. Beyond this age, males outperform females because maturation results in a shift in body composition. Our results are in line with previous investigations exploring physical capacities such as $\dot{V}O_{2\text{ peak}}$ and isometric strength in non-competitive or non-specialized adolescents [8–12,20].

Fig 3 shows that performance development during puberty varies considerably according to sport discipline. Within the jumping disciplines, improvement rates were slightly higher for long jump compared to high jump. Within the running disciplines, the improvement rates were slightly higher for 800 m compared to 60 m. Magnitudes of improvement within sex in jumping events were twice as large as in running events throughout the analyzed age stages. Moreover, tables 1 and 2 show that the peak rate of improvement in boys is reached at a later stage in jumping events (13–14 years) compared to running events (12–13 years). From a

motor learning perspective, running can be thought of as largely inherited, while jumping skills are to a larger degree affected by practice [21]. The relatively low performance progress in running exercises can possibly be explained, at least in part, by the fact that running is an innate movement typically developed naturally through growth and maturation from 1–2 years of age. A large part of the performance potential has thus already been exploited through natural play, growth and maturation. Jumping is also an innate movement, but not in the way in which long jump and high jump are executed in track and field athletics. These are performed with an approach speed from 7 to 12 m·s⁻¹, and with the athlete required to handle forces of up to 10 times their body weight [22,23]. It can therefore require many years of specific training to learn an optimal approach, take-off, flight- and landing technique. The relatively greater progress in jumping exercises can also be explained by growth and increased body height during puberty. The increase in body height means that the center of gravity will be higher, providing better mechanical conditions for performance in jumping events. Furthermore, Beunen & Malina [1] suggest that the plateau for some motor tasks occurs at a slightly older age. Previous studies of non-competitive adolescent boys indicate that peak gains in strength and power typically correspond with peak gains in body mass and muscle mass. Conversely, peak gains for running velocity and aerobic power typically occur at peak height velocity or slightly before [1,11,15]. The adolescent performance spurt for running speed is likely related to growth of the lower extremities, as the legs experience maximum growth before trunk, neck and head [24].

The current results indicate that the sex difference evolves from < 5% to 10–18% in all the analyzed disciplines from age 11 to 18 yr. The gap widens considerably during early adolescence before gradually stabilizing when approaching the age of 18. This evolution is practically identical for the running and jumping disciplines. The observed sex differences at the age of 18 are in line with previous studies of world-class athletes where a sex difference of 10–12% for running events [2,4,6] and ~19% for jumping events [6] has been reported. To the authors' knowledge, this is the first study to document when these sex differences arise.

A limitation of the present study is the lack of maturity status data. Youths who are successful in sport may differ in maturity status compared with the general population. Even at a later age, early maturing boys continue to have an advantage in strength and power tasks, while the later maturing boys typically catch-up when it comes to speed tasks [25]. In girls, maturity-associated variation in performance is not consistent among tasks and from age to age [9,26]. Most samples of female adolescent athletes have mean or median ages at menarche within the normal range, with the exception of gymnasts, ballet dancers, figure skaters and divers, who tend to be later than those of athletes in other sports [1]. In Belgian male track athletes, all 15–16 year old athletes except one had a skeletal age in advance of chronologic age, while in corresponding 17–18 years old athletes, two-thirds had skeletal age equal to or in advance of that expected for chronologic age [27]. However, it is worth noting that regular training does not affect growth, the timing and magnitude of peak height velocity, and skeletal and sexual maturation in young athletes [28] and the pubertal progress of boys and girls active in sport is similar to the progress observed in boys or girls not active in sport.

Another limitation is that only a handful of the present athletes were top 100 performers in every age category from 11 to 18 years of age. One could argue that early maturing athletes are overrepresented in the younger age stages, while they are successively replaced with average or late maturing athletes who catch up or outperform the early maturers when approaching late adolescence. Thus, the present data might overestimate individual annual performance development. However, falling trends among early maturers, at least up until a point where they are no longer within the top 100, are accounted for by the mixed models analysis. Furthermore, analysis of the small sample of athletes who were top 100 performers in every age category (2–6 athletes in each discipline) reveal similar performance development as our mixed model

analysis. When interpreting the current data, it should also be recognized that there are some limitations associated with using ratios and percentage change values [29]. Since the slope of the relationship between the logarithmically transformed numerator and denominator deviated somewhat from one for some variables, these ratios may not scale accurately towards the extremes of each range. However, mixed model analyses were performed on absolute rather than ratio values. Furthermore, percentage change and ratio data are presented supplemental to absolute values, and should therefore facilitate rather than mislead interpretation.

Because of the physical changes that occur during puberty, the optimal training program for this age group likely differs somewhat between boys and girls. For example, for adolescent girls, increases in fat mass and reductions in relative strength often occur alongside reductions in coordination and neuromuscular control [30]. This may impair training tolerance and increase the risk of certain types of injury [31]. Hence it may be beneficial for female athletes to have a greater focus on neuromuscular training during this period [32]. From a practical perspective, the present results allow coaches and athletes to set realistic goals and prescribe conditioning programs that take into account sex-specific differences in the rate of performance development at different stages of maturation. Because hormone-dependent changes during puberty mean that adolescent boys and girls respond somewhat differently to exercise, it is important that training is tailored based on sex and the specific stage of growth and maturation.

Author Contributions

Conceived and designed the experiments: ET TH AG. Performed the experiments: ET TH ISS. Analyzed the data: TH ISS ICO. Contributed reagents/materials/analysis tools: ICO ISS. Wrote the paper: ET TH ISS AG ICO.

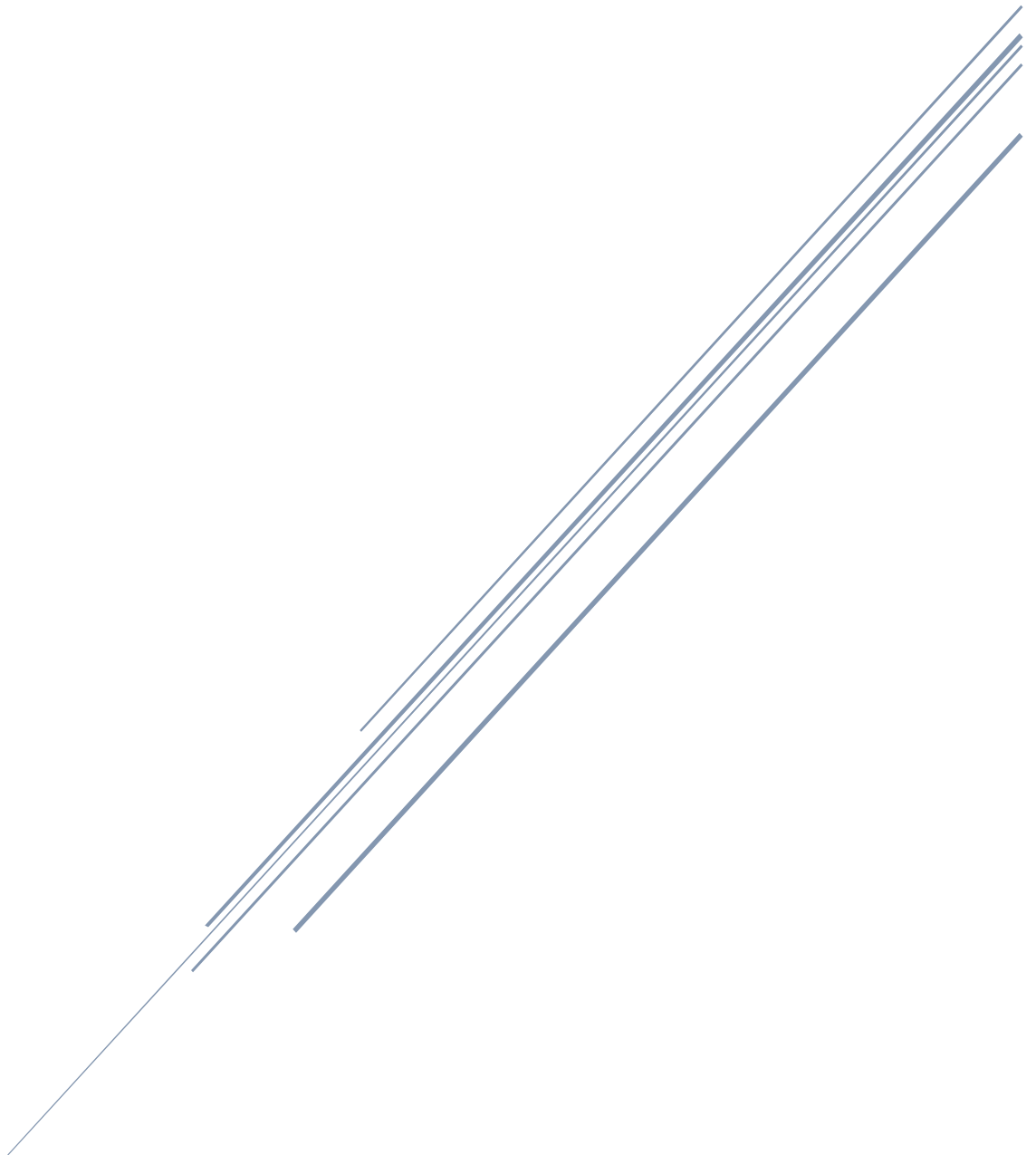
References

1. Beunen G, Malina RM. Growth, maturation and physical activity. In: Hebestreit H, and Bar-Or O, editors. *The young athlete*. Volume XIII of the encyclopaedia of sports medicine and IOC medical commission publication. Blackwell Publishing, 2008.
2. Cheuvront SN, Carter R, Deruisseau KC, Moffatt RJ. Running performance differences between men and women: an update. *Sports Med*. 2005; 35: 1017–24. PMID: [16336006](#)
3. Nevill AM, Whyte G. Are there limits to running world records? *Med Sci Sports Exerc*. 2005; 37: 1785–8. PMID: [16260981](#)
4. Seiler S, De Koning JJ, Foster C. The fall and rise of the gender difference in elite anaerobic performance 1952–2006. *Med Sci Sports Exerc*. 2007; 39: 534–40. PMID: [17473780](#)
5. Berthelot G, Thibault V, Tafflet M, Escolano S, El Helou N, Jouven X et al. The citius end: world records progression announces the completion of a brief ultra-physiological quest. *PLoS One*. 2008; 3:e1552. doi: [10.1371/journal.pone.0001552](#) PMID: [18253499](#)
6. Thibault V, Guillaume M, Berthelot G, Helou NE, Schaal K, Quinquis L, et al. Women and men in sport performance: The gender gap has not evolved since 1983. *J Sports Sci Med*. 2010; 9: 214–23 PMID: [24149688](#)
7. Tanner JM, Whitehouse RH, Marubini E, Resele LF. The adolescent growth spurt of boys and girls of the Harpenden growth study. *Ann Hum Biol*. 1976; 3: 109–126. PMID: [1275435](#)
8. Malina RM, Bouchard C, Beunen G. Human growth: Selected aspects of current research on well-nourished children. *Annual Review of Anthropology* 1988; 17:187–219.
9. Malina RM, Bouchard C, Bar-Or O. *Growth, maturation and physical activity*. 2nd ed. Champaign, IL: Human Kinetics; 2004.
10. Krahenbuhl GS, Skinner JS, Kohrt WM. Developmental aspects of maximal aerobic power in children. *Exerc Sport Sci Rev*. 1985; 13: 503–38. PMID: [3891374](#)
11. Armstrong N, Welsman JR, Nevill AM, Kirby BJ. Modeling growth and maturation changes in peak oxygen uptake in 11–13 yr olds. *J Appl Physiol*. 1999; 87: 2230–2236. PMID: [10601172](#)
12. Eisenmann JC, Malina RM. Age- and sex-associated variation in neuromuscular capacities of adolescent distance runners. *J Sports Sci*. 2003; 21: 551–557. PMID: [12848389](#)

13. Malina RM, Woynarowska B, Bielicki T, Beunen G, Eweld D, Geithner CA, et al. Prospective and retrospective longitudinal studies of the growth, maturation, and fitness of Polish youth active in sport. *Int J Sports Med.* 1997; 18: 179–185. PMID: [9187971](#)
14. Malina RM, Stawinska T, Ignasiak Z, Rozek K, Kochan K, Domaradzki J, Fugiel J. Sex differences in growth and performance of track and field athletes 11–15 years. *J Hum Kin.* 2010; 24: 79–85.
15. Geithner CA, Thomis MA, Eynde BV, Maes HHM, Loos RJF, Peeters M, Claessens ALM, Vlietinck R, Malina RM, Beunen GP. Growth in peak aerobic power during adolescence. *Med Sci Sports Exerc.* 2004; 36: 1616–1624. PMID: [15354046](#)
16. Goswami B, Roy AS, Dalui R, Bandyopadhyay A. Impact of pubertal growth on physical fitness. *Am J Sports Sci and Med.* 2014; 2: 34–39.
17. Siervogel RM, Demerath EW, Schubert C, Remsberg KE, Chumlea WC, Sun S, et al. Puberty and body composition. *Horm Res.* 2003; 60: 36–45. PMID: [12955016](#)
18. Norwegian Athletics Federation. All-time best statistics. Retrieved from: <http://www.friidrett.no/stevner/statistikk1/alle%20tidere/Sider/Norgesbestegjennomalletider.aspx> Updated Feb 1st, 2015.
19. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc.* 2009; 41: 3–13. doi: [10.1249/MSS.0b013e31818cb278](#) PMID: [19092709](#)
20. Valente-Dos-Santos J, Coelho-E-Silva MJ, Severino V, Duarte J, Martins RS, Figueiredo AJ et al. Longitudinal study of repeated sprint performance in youth soccer players of contrasting skeletal maturity status. *J Sports Sci Med.* 2012; 11: 371–9. PMID: [24149342](#)
21. Schmidt RA, Wrisberg CA. Motor learning and performance: A situation based learning approach, 4th edition. Human Kinetics; 2008.
22. Lees A, Graham-Smith P, Fowler N. A biomechanical analysis of the last stride, touchdown, and takeoff characteristics of the men's long jump. *J Appl Biomech.* 1994; 10: 61–61.
23. Seyfarth A, Friedrichs A, Wank V, Blickhan R. Dynamics of the long jump. *J Biomech.* 1999; 32: 1259–1267. PMID: [10569704](#)
24. Malina R. Top 10 research questions related to growth and maturation of relevance to physical activity, performance, and fitness. *Res Quart Exerc Sport.* 2014; 85: 157–173. PMID: [25098012](#)
25. Beunen G, Ostyn M, Renson R, Simons J, Van Gerven D. Motor performance as related to chronological age and maturation. In: Shepard RJ, Lavallee H, editors. *Physical fitness assessment: Principles, practice and application.* C.C Thomas Springfield, IL, 1978: 229–236.
26. Little NG, Day JAP, Steinke L. Relationship of physical performance to maturation in perimenarcheal girls. *Am J Hum Biol.* 1997; 9: 163–171.
27. Malina RM, Beunen G, Wellens R, Claessens A. Skeletal maturity and body size of teenage Belgian track and field athletes. *Ann Hum Biol.* 1986; 13: 331–339. PMID: [3767306](#)
28. Loucks AB, Vaitukaitis J, Cameron JL, Rogol AD, Skrinar G, Warren MP, et al. The reproductive system and exercise in women. *Med Sci Sports Exerc.* 1992; 24: 288–289.
29. Atkinson G, Batterham AM. The use of ratios and percentage changes in sports medicine: time for a re-think? *Int J Sports Med.* 2012; 33: 505–506. doi: [10.1055/s-0032-1316355](#) PMID: [22760546](#)
30. Hewett TE, Myer GD, Ford KR. Decrease in neuromuscular control about the knee with maturation in female athletes. *J Bone Joint Surg Am.* 2004; 86-A: 1601–1608. PMID: [15292405](#)
31. Tursz A, Crost M. Sports-related injuries in children. A study of their characteristics, frequency, and severity, with comparison to other types of accidental injuries. *Am J Sports Med.* 1986; 14: 294–299. PMID: [3728781](#)
32. Myer GD, Ford KR, Hewett TE. Methodological approaches and rationale for training to prevent anterior cruciate ligament injuries in female athletes. *Scand J Med Sci Sports.* 2004; 14: 275–285. PMID: [15387801](#)

INTERNATIONAL RESEARCH LITERATURE REVIEW

SCEG Project for Review and Redraft of Guidance for
Transgender Inclusion in Domestic Sport 2020



Carbmill Consulting

DEFENDANT'S
EXHIBIT

23

Carbmill Consulting

This document is produced specifically as part of the Review and Redraft of the Guidance for Transgender Inclusion in Sport for the Sports Councils' Equality Group of UK.

The literature review includes published scientific research relating to transgender inclusion in sport, and is based largely on peer-reviewed data, but also some pertinent opinion pieces (outlined in the text). It is written in a general chronological order, with reference to important milestones in international sport to provide context and relevance to historical events and policy decisions. In keeping with the remit from UKSport, the document focusses on scientific and medical literature related to performance, rather than sociological or ethical considerations. For simplicity and clarity, the terminology used here is male/female and transgender man/transgender woman/transgender people.

Contents

Executive summary	1
Introduction	2
Population-based differences between the sexes.....	2
Derivation of performance difference between the sexes in sport	4
History of sex categories and inclusion of transgender athletes.....	4
Debate about the Role of Testosterone	8
Medical and Psychosocial Aspects.....	8
Ethical Considerations.....	10
Conclusion.....	10
REFERENCES	11

Executive summary

The research base for the inclusion of transgender people in sport remains limited but is rapidly evolving.

The evidence outlining the sex dimorphism, or physical variance, between the sexes is robust: There are significant differences between the sexes which render direct competition between males and females unfair in most 'gender-affected sports' due to the physical advantages of males, and often unsafe in sports which allow physical contact and collisions.

Recent international guidelines, and subsequent policies, on the inclusion of transgender people (particularly transgender women) have attempted to negate male physical advantage, thereby allowing fair (and safe) competition. However, there is currently no direct evidence that this can be achieved by suppression of hormone levels. On the contrary, there are apparent life-long physiological advantages in the adult male, only some of which can be reversed, and it is not known whether physical benefit(s) may be retained in someone who does not achieve full maturity.

Transgender people face health concerns related to their transition and are prone to chronic side effects which are likely to be ameliorated by physical activity. On the other hand, female athletes are subject to specific health and injury concerns in direct relation to their reproductive physiology and

Carbmill Consulting

cyclic metabolism. It is not known if these differences create a meaningful disparity in performance sport.

Everyone benefits from physical activity for both physical and mental health, and there are reported advantages in relation to the social aspects of team sports in particular. These benefits should be available to everyone in our community, and it is known that transgender people face barriers in access to sport.

This short review presents a chronological precis of current evidence which impacts decision making on how best to include transgender people in sport.

Introduction

Transgender inclusion transects the established categorisation in sport and, by including those who have transitioned from the alternate sex category, has created challenges. The binary of sex categories in sport has existed for more than a century and arguably since the beginning of sport in ancient times when females were largely excluded. Sports have historically been developed for able-bodied males by able-bodied males, and the opportunity for females, children, masters' athletes, and disabled people has been an evolution of increased opportunity and inclusion through categorisation over many decades. It is important to note that all subsequent categories within competitive sport, including weight categories, have always been underpinned by the sex binary, and inclusion of other groups has been largely by the addition of new and separate competition within that binary.

Research literature for the evidential basis of inclusion for transgender people in sport remains limited. The definitive longitudinal studies tracking prospective gender transition in athletes, with appropriate objective measurements, competition results and with a relevant control population, have not yet been reported. Further, the scientific literature on performance differences between males and females across the sports is also limited because, until now, the relevance of this difference was questionable as the sexes have not routinely competed against each other. Hence much of the known information is laboratory based, rather than specific to sporting competition.

Population-based differences between the sexes – the rationale for gender-affected competition

In adulthood, there are clear differences in muscle mass, strength, and physiological capacity even after controlling for differences in height and weight between the sexes. On average, males have 40-50% greater upper limb strength, 20-40% greater lower limb strength, and an average of 12kg more skeletal muscle mass than age-matched females at any given body weight (Janssen et al 2000, Handelsman et al 2018).

Following full growth, males are measured at an average 13-15cm (5-6 inches) taller than females. The National Health Service in 2018 quotes the average young adult male in England at a height of 178.0cm (5ft 10in) with an estimated standard deviation (SD) of 9.19cm. The average young adult female in England is 164.1cm (5ft 4 1/2in), SD 7.51cm. In the USA the statistics are very similar: Males average 175.9cm, SD 7.42cm, and females average 162.1cm, SD 7.11cm.

Carbmill Consulting

When plotted across the population it is evident that 50% of males are taller than some 97% of females. So, in any random grouping of males and females, half of the males will be taller than nearly all females.

Handgrip dynamometry is probably the most commonly used method to characterize overall human muscle strength (Bohannon et al 2019). European data of over 2000 young adult males and females demonstrates “mean maximal hand-grip strength showed clear difference between men (541 Newtons) and women (329 Newtons); 90% of females produced less force than 95% of males”. And further.... “highly trained female athletes corresponded to only the 25th percentile of the (untrained) male subjects” (Leyk et al, 2007).

Therefore, this would mean that in a random group of ten males and ten females, only the strongest female is likely to be stronger than the weakest male.

Males have a larger, stronger and denser skeleton, with longer levers (arms and legs), and bigger hands and feet. Males also have larger hearts and greater blood volume with higher haemoglobin concentration, and hence greater oxygen carrying capacity, with bigger lungs and airway capacity compared with females. The differences in physical/physiological capacity are synergistic such that males are larger, stronger and faster, with greater aerobic capacity and for which the summation is likely to be greater than the individual differences.

The muscle strength, physiological and size differences between the sexes correlate with significant sporting achievement from around 10-12% for most linear swimming and running events, 20% difference in jumping events, and a 35% difference in weightlifting ability in weight matched males and females. If an average sized male is compared with an average size female (based on the NHS data, 2018) the weightlifting ability increases to 50% greater for males (World Record Snatch; 85kg Male - 187kg lift; 69kg Female - 123kg lift).

It is often assumed that children have similar physical capacity regardless of their sex, and hence mixed competition is allowed in many sports up until the age of around 11 years (to coincide with puberty). However large-scale data reports on children from the age of six show that young males have significant advantage in cardiovascular endurance, muscular strength, muscular endurance, speed/agility and power tests, but score lower on flexibility tests (Tambalis et al, 2016; Catley & Tomkinson, 2013).

In adults the average advantage of 10-12% in running events can be interpreted for both the ‘average’ competitor, as well as for elite performers where the percentage difference remains the same for similarly trained people. When measured in absolute time, i.e. the difference to reach the finish line, this will vary based on the ability of the athletes; but the distance between the competitors remains the same. In this way a 10-12% better performance by a competitor would see Adam Peaty being beaten by half a length in the short-course 100m breaststroke and Dina Asher Smith by 22m in the 200m track sprint. And Sir Mo Farah would be lapped twice in the 10,000m track race.

While there may be many differences between competitors of the same sex (such as height, skill, and training methods), what is relevant about the advantage between the sexes and the reason for the category division, is the scale of advantage across virtually all sports. World Rugby (2020) has referred to “...the almost perfect sensitivity of biological sex.... since in a ranking list of the top thousand performances in most sport, every year, every single one will be biologically male”. The winning margin (the difference in performance by which a competitor misses a gold medal, any medal, or making the final) in elite athletics or swimming events during the last three Olympics is

Carbmill Consulting

less than just 1% for both male and female events, compared with up to a 50% difference between sex-based performances.

Derivation of performance difference between the sexes in sport

The known dichotomy in athletic performance between the sexes has long been acknowledged to largely be in relation to the difference in gonadal androgens, namely testes derived testosterone and its more potent metabolite dihydrotestosterone in males, and ovarian estrogen in females.

Males are prenatally (before birth) and perinatally (soon after birth) exposed to high level androgens, and then experience a further surge in testosterone levels during puberty, which is maintained throughout adult life. On the other hand, females normally never experience testosterone exposure beyond a basal level at any time during life ($<2\text{nmol/L}$). From the onset of male puberty, the testes secrete 20 times more testosterone, resulting in circulating testosterone levels that are 15 times greater in healthy young males than in age-similar healthy females (Handelsman et al 2018).

However, females enter maturational puberty at least a year before males, and this may create a brief equivalence in the height of children around the age of 10, before males enter puberty from around the age of 11.

Anabolic steroids, of which testosterone is the significant 'natural' form, were one of the original groups of banned drugs when anti-doping practices developed in the 1970s. Both endogenous (made within the body) and exogenous testosterone (delivered into the body from an external source) are identical chemical substances and have equivalent effects in the body. Testosterone concentration positively correlates with leg press strength, thigh muscle volume and haemoglobin levels, and the anabolic response to testosterone can largely be predicted by the dose administered (Bhasin et al 2001). Testosterone is largely responsible for the increased muscle mass through increases in muscle fibre numbers and size, muscle satellite cell numbers, numbers of myonuclei (muscle cells nucleus), and size of motor neurones (nerve cells) as well as advantageous biochemical changes in muscle function in males. Importantly stimulation of myonuclei is life-long (Handelsman et al 2018). This is so-called 'muscle memory' and is one of the reasons why doping sanctions were increased to four years to try to account for this long-term effect on performance.

For ethical and doping control reasons it has not been possible to evaluate the effect of testosterone in healthy young athletes beyond cross sectional studies, but observational reports have lent credence to this understanding. During the era of the former German Democratic Republic, systematic doping of athletes was developed and later confirmed on release of the so-called 'Stasi Files'. Regime scientists stated after the 1972 Munich Olympics: "the effects of the treatment with androgenic hormones were so spectacular, particularly in female athletes in strength dependent events, that few competitors not using the drugs had a chance of winning". Further, the long-term effect of androgenic/anabolic steroids was attested by the acknowledgement that "after a critical period of increase in muscle strength, a higher performance level is reached that does not return to pre-treatment values after the drug is withdrawn" (Franke & Berendonk 1997).

History of sex categories and inclusion of transgender athletes

Previous practices of 'gender (sex) verification' for female athletes were developed "in order to exclude those in which a genetic alteration gives the competitor registered as a female anatomical advantages of masculinization" (Hay 1985). There was no perception that females would pose a

Carbmill Consulting

competitive threat to male athletes. Genetic testing, which received “great support” from athletes, was contentious within the scientific community and was subsequently abandoned in 2000 for the Sydney Olympics (Ljungqvist 2000). This was largely due to some infamous cases in which inadequate processes resulted in inappropriate disqualification of athletes.

The Stockholm Consensus of the IOC in 2003 established the first international policy for transgender inclusion in international sport. This was not a referenced document; “at the time, there was no published scientific literature that would justify the inclusion of transgender women” (Harper 2015). Transgender people were at this time required to have full surgical transition in order to compete in their acquired gender.

Dr Arne Ljungqvist, the then IOC Medical Commission Chair, is reported (Harper 2015) to have relied heavily on the opinion of Dr Louis Gooren who went on to publish in 2004 the first comparative study of the physical/physiological attributes of transsexual men and women. In this study the metabolic and morphological changes of 17 transgender men and 19 transgender women were studied before and after their gender-affirming medication. The pre-treatment data serves as control groups for the respective treatment groups for statistical analysis. Muscle mass (measured by MRI) and serum haemoglobin were compared across one and then three years of treatment.

Following one year of androgen(testosterone) deprivation, muscle mass decreased significantly in the transgender women, but remained significantly higher than the pre-treatment transgender men (female) group. On the other hand, the transgender men on testosterone supplementation increased muscle mass such that there was no statistical difference between these subjects and the pre-treatment transgender women (male) group. There was no difference between the treatment groups (transgender men and transgender women) after one year of treatment. There was a further non-significant loss of muscle mass following a further two years of treatment of transgender women. Haemoglobin levels converted into the respective ranges of the acquired gender in each group within the first year. This would have a corresponding effect on blood oxygen carrying capacity, and hence aerobic ability. It was noted that height was a strong independent predictor of muscle mass in both males and females, and both pre- and post- gender transition (Gooren and Bunck 2004).

Following cessation of gender/sex (chromosomal) testing in 2000, athletes with Disorders (Differences) of Sexual Development (DSD) were successful in major athletic competitions; notably at the Athletics World Championships in 2009. There followed a decade during which several athletes with the genetic disorder ‘46XY 5ARD’ (46 XY chromosome women in which an enzyme abnormality causes testosterone to rise to high levels) were in dispute with the IAAF/World Athletics through lengthy court proceedings in relation to their hyperandrogenic status. While these cases do not have direct relevance to transgender inclusion in sport what is important is that as a result of the requirement for evidence and information for court documentation, there was a significant increase in both the research and academic literature on the effect of testosterone in sports performance over this period.

During the Daegu World Athletics Championships in 2011, research on behalf of the IAAF and WADA on the normative values of serum androgen levels of elite track and field athletes was conducted in order to inform the blood steroidal module of the Athlete Biological Passport for doping control (Bermon et al 2014). This research was able to define the median value of serum testosterone in elite female athletes at 0.69 nmol/L with a 25-75% range of 0.50 to 0.91nmol/L, and a 99th percentile calculated at 3.08nmol/L. The testosterone levels in the DSD group were as high as 29.30nmol/L, while the standard male range of testosterone is 7.7 to 29.4nmol/L (Handelsman 2018). It was

Carbmill Consulting

calculated that, within the elite athlete population, hyperandrogenism in relation to XY DSD was 140 times more common than in the general population.

Researcher Joanna Harper published comparative data in 2015, around the time that the second IOC Consensus Statement on Sex Reassignment was published. In this study, the race times of eight transgender/transsexual women were recorded both pre- and post- transition. All subjects were recreational distance runners who self-reported race times (which were verified where possible) in equivalent events over a period stretching from one year up to twenty-seven years apart. No testosterone levels were reported for the athletes, either before or after the transition. Calculations were made to compare their race times with those of age-graded performances in the respective gender. The findings indicated that the comparative race times were equivalent pre- and post-transition, and this implies that transition did not give these transgender women a competitive advantage (Harper 2015). This was the first study to attempt to measure physical performance following transition, compared with pre-transition ability and, while there were several acknowledged limitations to the study, this information appears to have been crucial in the subsequent deliberations of the second IOC Consensus Meeting.

The second Consensus Paper of the IOC was published in November 2015. Surgery is no longer required to be eligible to compete, and transgender women are required to suppress testosterone levels to below 10nmol/L for 12 months before competition. This testosterone level is 5-10 times the testosterone level of females of any age and remains within the accepted range for males, and it has not been defined as to how this level was determined by the IOC. Transgender men can enter male competition without requirement.

Over the past two decades there have been a dozen published research papers (five of which were published in 2018-20) for a total of some 800 participants which investigated the effect of testosterone suppression on muscle mass, and in some instances muscle strength, in transgender women.

Post-suppression testosterone levels measured in these studies were all less than 10nmol/L, most being below 5nmol/L, and seven studies where testosterone level was less than 1nmol/L (ie at the level of females). Most studies included estrogen therapy for participants. On review by Hilton and Lundberg, the findings from the composite of existing research was that loss of muscle mass and/or strength following 12 months of testosterone suppression is relatively modest compared with the initial advantage. The average loss of muscle mass across all studies was around 5%, which is less than the 9% loss measured in the seminal work of Gooren and Bunck. In those studies which recorded the retained muscle mass/strength, there was an average of 25% residual advantage for transgender women at 12 months treatment compared with reference a group of females (Hilton and Lundberg 2020).

In the most recent study, by Wiik and colleagues from the Karolinska Institute in Sweden, there was minimal loss of muscle mass or strength in transgender women following 12 months of testosterone suppression, and some subjects gained strength. After treatment transgender women remained 48% stronger, with 35% larger quadriceps mass compared with the control population of females (Wiik et al 2020).

Harper and colleagues also reviewed the available scientific literature and published in the British Journal of Sports Medicine in March 2021. They concluded that “haemoglobin levels rapidly reduce in transwomen” while “strength may be well preserved during the first 3 years of hormone therapy” (Harper et al 2021).

Carbmill Consulting

Recent research citing retrospective data from military personnel in the US has shown that transgender women retain an advantage in running speed after more than two years of follow-up on testosterone suppression, at a residual of some 12% faster than the known normative values for females. In the same study, values for numbers of 'push-ups' and 'sit-ups' had reverted to female levels, but data on training levels and performance targets was lacking. Results for the transgender men studied exceeded those of the average performance for age-matched males after one year of testosterone supplementation (Roberts et al, 2020).

From the synthesis of current research, the understanding is that testosterone suppression for the mandated one year before competition will result in little or no change to the anatomical differences between the sexes, and a more complete reversal of some acute phase metabolic pathways such as haemoglobin levels although the impact on running performance appears limited, and a modest change in muscle mass and strength: The average of around 5% loss of muscle mass and strength will not reverse the average 40-50% difference in strength that typically exists between the two sexes. These findings are at odds with the accepted intention of current policy in sport, in which twelve months of testosterone suppression is expected to create equivalence between transgender women and females.

It is recognised that research to date has not been carried out in an elite athletic population (although military data includes well-trained individuals), and it is not known if the relative muscle loss will be more or less in those who are undergoing training, relative to the published results in non-athletic populations. It is known however, that males who undergo androgen suppression as part of treatment for prostate cancer can improve muscle strength and mass when also undertaking a training program, whereas the absence of training results in 4% reductions in muscle mass, similar to those observed in trans women whose testosterone is suppressed without training (Galvao et al 2006). The implication is that training during a period of testosterone suppression may attenuate the reductions observed in the longitudinal studies, which would increase the retained advantage when comparisons are made with a reference group of females.

The notion of safety in sport is related to size and maturity of competitors and is one of the reasons that junior athletes are not routinely eligible to play against adults in contact and collision sports. Because the two sexes have not historically competed against each other as adults in these sports, there is not yet experience of the injury risk to which females may be exposed in such competition, nor with the inclusion of transgender people beyond anecdotal reports. Injury risk can be appreciated through analysis of the combination of mass and speed of competitors and hence the force that can be generated, in which the larger and faster opponent will pose a risk to those that are smaller and slower when tackled. At a conference convened by World Rugby in London in February 2020, information was presented (and subsequently published as part of World Rugby Transgender Guideline) of a theoretical framework for injury risk based on validated mechanical modelling, relevant to the size differential of male players compared with female players. While not including the speed of a bigger opponent, nor the neck strength of the smaller, the increased mass alone accounted for an estimated increase in the risk factors for head and neck injury of 20-30%. Given the modest reduction of body mass by testosterone reduction in transgender women this was estimated to pose an unacceptable risk to female players and, in part, resulted in the decision to render transgender women ineligible to compete in international contact rugby union (World Rugby 2020).

Carbmill Consulting

Debate about the role of testosterone

Writing an 'Observation' in the Scientific American journal in June 2019, Rebecca Jordan-Young and Katrina Karkazis, authors of "Testosterone; an Unauthorised Biography", question the validity of ascribing the differences between male and female athletes solely to endogenous levels of testosterone. They draw attention to the inconsistency between measured testosterone levels and performance. The research published on behalf of the IAAF (Bermón & Garnier 2017) received criticism from the scientific community and required formal correction. Jordan-Young and Karkazis stress that in three of the eleven running events cited, it was the group with the lowest testosterone levels who performed best, and there was a strong negative association between testosterone levels and 100m race time.

The authors quote published work from Pielke, Tucker and Boye (2019), which questions the validity of IAAF(WA) research which was requisite by the Court of Arbitration in Sport (CAS) to demonstrate that "high androgen levels confer a significant performance advantage over other female competitors, comparable to the performance advantage that male athletes enjoy over female athletes".

In the event, the full evidence base presented by IAAF/World Athletics was ultimately successful at CAS, and this was defended again on appeal. However, the contention remains in the scientific community that the case has not been made that circulating testosterone can fully explain all the observed difference in male/female performance variables (Jordan-Young & Karkazis 2019).

Importantly, the corollary of circulating testosterone levels not being fully accountable for this difference is that there must be another, as yet undefined factor(s), which must account for the remaining differences following testosterone suppression. The retained advantage may mean that previous testosterone exposure is responsible for long-term effects, or there is another explanation, perhaps referable to genetic differences, which has not yet been identified.

Following the legal cases, World Athletics also created policy for the inclusion of transgender women based on suppression of testosterone levels to 5nmol/L. Between 2018 to 2019, several major sporting international governing bodies developed policies at the 5nmol/L level for transgender women, in comparison with the 10nmol/L defined by the IOC.

Medical and Psychosocial Aspects

A review published in 2017 by Bethany Jones and colleagues from Nottingham/Loughborough Universities looked at research that explored the experience of transgender people in sport, as well as the policies which were in place at the time which influenced such experience. While only eight research papers met the criteria for the study, in those which looked at the experience of transgender inclusion, it was acknowledged that the experience of transgender people in sport was largely negative. The authors' assessment of the available literature on physical parameters was to highlight the lack of direct comparative data in athletes on which to base exclusionary policies. They indicate that 'There are several areas of future research required to significantly improve our knowledge of transgender people's experience in sport, inform the development of more inclusive sport policies, and most importantly, enhance the lives of transgender people, both physically and psychosocially' (Jones et al, 2017).

Differences in physiology and health parameters between males and females have long been acknowledged and recent research has shed light on how this affects performance, as well as health,

Carbmill Consulting

illness and injury. The menstrual cycle creates challenges for many female athletes and the variable nature of ovarian hormones can limit training and performance, and also alter perceptions of wellbeing. Furthermore, amenorrhoea (loss of menstrual cycling) is common in sport due to high training loads and inadequate energy intake. This has secondary consequence of not just altered fertility but increased rates of injury and illness. The naturally lower bone density of females is further reduced with the hormonal deprivation of amenorrhoea and this group is particularly prone to stress fractures, a significant problem in running sports. Female athletes are not immune to gynaecological illnesses, and endometriosis which has an incidence of up to 20% in the general population, has plagued the career of several high-profile athletes who have chosen to go public with their inability to compete as a result of pain and illness.

Recent research has highlighted the relationship of the menstrual cycle and the observed increased injury incidence and slower recovery in women compared with men, in particular for concussion (McGroarty et al 2020), and major ligament injury such as anterior cruciate ligament tears (Wojtys et al 2002). The physical problems related to the menstrual cycle are now so well appreciated that ongoing research in health and hormonal monitoring is taking place in the UK (Football Association) and abroad. This research is designed to define parameters whereby limitations to training and performance can be minimised.

By way of comparison, it is known that transgender people suffer increased risk of ill-health as a result of medication and cross-sex hormone therapy. Transgender women face heightened risk of thrombo-embolic disease, arterial disease, liver dysfunction and breast cancer (although at levels lower than females). Transgender men risk high red cell levels (polycythaemia), hypertension and arterial disease, as well as reproductive cancer. These are chronic conditions for which exercise is both preventive and therapeutic and are unlikely to present any direct negative effects in relation to sporting performance. Certainly, evidence indicates that bone health and density is maintained in transgender women despite testosterone suppression (Van Caenegem et al 2015).

Over time it may become apparent that medically transitioning will lead to health and injury issues specific to transgender athletes, but this will not have the cyclical nature of gynaecological function and any sport specific concerns are likely to be different to those experienced by females and males. It will be important to track and define the health consequence of the required medical treatments required by sports agencies to compete in one's affirmed gender.

We are not aware of research which will allow any conclusions to be drawn on the current medical practice of transition in younger people, and what effect this may have on physical and sporting capacity. In the UK, so-called 'puberty blockers' are generally not used until Tanner maturation stage 2-3 (i.e. after puberty has progressed into early sexual maturation), and medical practice is currently under review. As described above it is evident that current knowledge cannot explain the exact derivation of all the performance differences between the sexes, and correspondingly what, if anything, can be done to reverse these differences in adults.

While exercise and activity are known to positively impact mental health, elite level training and competition are known to increase stress. Both competitive athletes and transgender people have separately reported high levels of mental ill-health, notably anxiety states, depression, and poor coping mechanisms with stress: These include self-harm and eating disorders. It is not known how the stresses of competition may affect transgender athletes.

Reports of suicidality in transgender people are contentious. A recent large-scale meta-analysis of five peer-reviewed papers of over 1000 study subjects found a prevalence of suicide attempts of 14.8% (Surace et al, 2020). This is at odds with other commissioned reports of smaller sample sizes which give much higher estimates.

Carbmill Consulting

There is no contention however as to the value of physical activity and community membership on mental health outcomes. These benefits should be available to all.

Ethical Considerations

The notion of ethics within sport is largely based on established theories of fairness and fair play, advantage, inclusion, safety and values. The competing 'rights' of individuals can be explored through both legal and philosophical discourse. These writings are beyond the scope of this document; however, it is relevant to consider the implication of sporting rules (such as testosterone suppression or use of puberty blockers) which may or may not have a desired outcome, and which have can have unintended poor health outcomes for the individual. Further, the development of protocols within sport which create an environment whereby 'case-by-case' analysis judgements of an individual may restrict their rights, as well as those of others, can be fraught.

Conclusion

Sexual dimorphism in relation to sport is significant and the most important determinant of sporting capacity. The challenge to sporting bodies is most evident in the inclusion of transgender people in female sport.

Current international rules require testosterone suppression to achieve equity in female sporting competition. In the light of existing evidence, the following is likely to be true:

Testosterone supplementation in transgender men results in strength and stamina (haemoglobin) parity with males within twelve months. Skeletal differences (smaller physique) will remain in keeping with female dimensions.

Testosterone suppression has differing effects for different body systems, and this can be correlated with the factors of 'gender-affected sport'; Strength, Stamina and Physique;

- Strength; Modest change within 12 months: Muscle mass (and perhaps cardiac size) and hence strength - appears retained at significantly higher levels than females.
- Stamina; Restoration of haemoglobin levels to female typical levels within 12 months: with relevant effect on oxygen carrying capacity as yet undefined.
- Physique; Minimal change: Structural features including the skeleton, bone density, height, lung and airway size, and tendon/ligament strength will remain, with modest loss of muscle mass.

This evidence suggests that parity in physical performance in relation to gender-affected sport cannot be achieved for transgender people in female sport through testosterone suppression. Theoretical estimation in contact and collision sport indicate injury risk is likely to be increased for female competitors.

There are significant ethical and legal considerations associated with the requirement for treatment which may not result in the desired outcome. Ongoing research may allow better understanding of comparative data to inform decision making.

Carbmill Consulting

The health and social imperative to include transgender people in physical activity and sport is paramount. The task for the National Governing Bodies of sport is to achieve safe, fair and inclusive sport, and/or inclusive, fair and safe sport. How this can be achieved will be explored within this review process.

REFERENCES

- Bermon S, Garnier PY, Hirschberg AL, et al (2014). Serum Androgen Levels in Elite Female Athletes; *The Journal of Clinical Endocrinology & Metabolism*; Vol 99: Issue 11, 4328-4335
- Bermon S, Garnier PY (2017). Serum androgen levels and their relation to performance in track and field: mass spectrometry results from 2127 observations in male and female elite athletes. *British Journal Sports Medicine*; Vol 51: Issue 17
- Bhasin S, Woodhouse L, Casaburi R, Singh AB et al (2001). Testosterone dose-response relationships in healthy young men. *American Journal of Physiology, Endocrinology and Metabolism*; Vol 281: E1, 172-181.
- Bohannon RW, Wang YC, Yen SC, Grogan KA (2019). Handgrip strength: A Comparison of Values Obtained from the NHANES and NIH Toolbox studies. *American Journal of Occupational Therapy*; March/April, Vol 73: No 2, 1-9
- Catley MJ & Tomkinson GR (2013). Normative health-related fitness values for children: analysis of 85347 test results on 9-17-year-old Australians since 1985. *British Journal of Sports Medicine*; Vol 47: 98-109.
- Equality Act (2010) UK <https://www.legislation.gov.uk/ukpga/2010/15/contents>
- Franke WW, Berendonk B (1997). Hormonal doping and androgenization of athletes: a secret program of the German Democratic Republic Government. *Clinical Chemistry*; Vol 43(7):1262-1279
- Galvao DA, Nosaka K, Taaffe DR et al (2006). Resistance training and reduction of treatment side effects in prostate cancer patients. *Med Sci Sports Exercise*; 38:2045-52.
- Gooren LJG and Bunck MCM (2004). Transsexuals and competitive sports. *European Journal of Endocrinology*; 151: 425-429
- Handelsman DJ, Hirschberg AL, Bermon S; (2018). Circulating testosterone as the Hormonal Basis of Sex Differences in Athletic Performance: *Endocrine Reviews*; Vol 39: 803-82
- Harper, J (2015). Race Times for Transgender Athletes: *Journal of Sporting Cultures and Identities*; Vol 6: Issue 1
- Harper J, O'Donnell E, Khorashad BS, McDermott H, Witcomb GL (2021). How does hormone transition in transgender women change body composition, muscle strength and haemoglobin? Systematic review with a focus on the implications for sport participation. *British Journal of Sports Medicine*; 10.1136/bjsports-2020-103106
- Hay E (1985). Femininity Controls in the Olympic Games: Sports Medicine in Track & Field Athletics; *IAAF Medical Congress (Proceedings)*; Canberra, Australia, pages 49-52.

Carbmill Consulting

Hilton EN, Lundberg TR (2020). Transgender Women in the Female Category of Sport: Perspectives on Testosterone Suppression and Performance Advantage. *Sports Medicine*; <https://doi.org/10.1007/s40279-020-01389-3>

Janssen I, Heymsfield SB, Wang ZM, Ross R (2000). Skeletal muscle mass and distribution in 468 men and women aged 18-88yr. *J Appl Physiol*; Vol 89(1):81-88

Jones, BA, Arcelus J, Bouman WP, Haycraft E. (2017) Sport and Transgender People: A Systematic Review of the Literature Relating to Sport Participation and Competitive Sport Policies. *Sports Medicine*; 47: 701-716

Jordan-Young RM, Karkazis K. (2019) 4 Myths about Testosterone. *Scientific American*; 18 June 2019. <https://blogs.scientificamerican.com/observations/4-myths-about-testosterone/>

Leyk, D, Gorges, W, Ridder, D et al. (2007). Hand-grip strength of young men, women and highly trained female athletes. *European Journal of Applied Physiology*; 99: 415–421. <https://doi.org/10.1007/s00421-006-0351-1>

Ljungqvist A (2000). Gender Verification: *The Encyclopaedia of Sports Medicine (IOC & ISMF)*; Vol VIII: Chapter 13

McGroarty NK, Brown SM, Mulcahey MK (2020). Sport-Related Concussion in Female Athletes, A Systemic Review: *Orthopaedic Journal of Sports Medicine*; 8(7) [https://DOI: 10.1177/2325967120932306](https://DOI:10.1177/2325967120932306)

National Health Service: Health Survey for England (2018) <http://digital.nhs.uk/pubs/hse2018>

Pielke, R, Tucker, R & Boye, E (2019). Scientific integrity and the IAAF testosterone regulations. *International Sports Law Journal*; 19: 18–26. <https://doi.org/10.1007/s40318-019-00143-w>

Roberts TA, Smalley J, Ahrendt K (2020). Effect of gender affirming hormones on athletics performance in transwomen and transmen: implication for sporting organisations and legislators. *British Journal of Sports Medicine*. Accepted 28 August 2020. <http://dx.doi.org/10.1136/bjsports-2020-102329>

Surace T, Fuser-Poli L, Voza L, Cavone V, Arcidiacono C et al (2020). Lifetime prevalence of suicidal ideation and suicidal behaviors in gender non-confirming youths: a meta-analysis. *European Journal Child Adolescent Psychiatry* Mar 13, DOI: [10.1007/s00787-020-01508-5](https://doi.org/10.1007/s00787-020-01508-5)

Tambalis KD, Panagiotakos DB, Psarra G, Daskalakis S, Kavouras SA, Geladas N, Tokmakidis S & Sidossis LS (2016). Physical fitness normative values for 6-18-year-old Greek boys and girls, using the empirical distribution and lambda, mu, and sigma statistical method. *European Journal of Sport Science*; 16:6 736-746

Van Caenegem E, Wierckx K, Taes Y et al (2015). Preservation of volumetric bone density and geometry in transwomen during cross-sex hormonal therapy: a prospective observational study. *Osteoporosis Int*; 26(1): 35-47

Wiik A, Lundberg TR, Rullman E et al (2020). Muscle strength, size, and composition following twelve months of gender-affirming treatment in transgender individuals. *Journ Clin Endocrinol Metab*; 105(3):805-813 doi:10.1210/clinem/dgz247

Carbmill Consulting

Wojtys EM, Huston LJ, Boynton MD, Spindler KP& Lindendfeld TN (2002). The Effect of the Menstrual Cycle on Anterior Cruciate Ligament Injuries in Women as Determined by Hormone Levels; *American Journal of Sport Medicine*; 30(2): 182-188.

World Rugby: Player Welfare and Guidelines: 2020

<https://playerwelfare.worldrugby.org/?subsection=84>

<https://playerwelfare.worldrugby.org/?subsection=5>

The ‘Motorik-Modul’ (MoMo): physical fitness and physical activity in German children and adolescents

Alexander Woll · Bärbel-Maria Kurth · Elke Oppen ·
Annette Worth · Klaus Bös

Received: 20 July 2010 / Accepted: 23 December 2010 / Published online: 12 February 2011
© Springer-Verlag 2011

Abstract

Objectives The aim of the ‘Motorik-Modul’ (MoMo Basiserhebung, www.motorik-modul.de) was to establish prevalence measurements on physical fitness and physical activity in German children and adolescents and to identify differences between age groups and genders. A total of 4,529 children and adolescents between the ages of 4–17 years from 167 cities across all states of the German Federation participated in this study. Sociodemographic parameters were recorded. Physical fitness was measured using endurance, strength, coordination and flexibility tests. Physical activity was assessed using a validated questionnaire, including aspects of physical activity participation (school/club/non-club), intensity of physical activity and type of activity. Chi-square tests and ANOVAS were used to compare groups and to detect age and gender effects, and data were compared with criterion-related guidelines by the WHO. For parameters describing physical fitness, age was the dominant influencing parameter. Many parameters showed increasing performance until the onset of puberty. Boys generally performed better than girls except in fine motor coordination during precision tasks and in flexibility. Physical

activity was greater in older subjects with a greater intensity for adolescents who were physically active. The general activity level was similar to the data reported in the HBSC study by the WHO but lower than the recommended activity levels.

Conclusion With the results of this study, nationwide representative prevalence data on physical fitness and physical activity are available. These data contribute to the scientific knowledge on the prevalence and conditions of physical fitness and physical activity.

Keywords Physical fitness · Motor performance ability · Sports · Physical activity · Age · Gender

Introduction

Physical fitness and physical activity are important determinants of health. Especially in children and adolescents, physical activity is a central component of health behaviour, and physical fitness is often used to evaluate a child’s general health and motor development. While a sedentary lifestyle and lack of physical activity have profound health implications for persons of all ages, they may hinder normal development and growth in children and adolescents and result in life-long debilitated health. Negative correlates of lack or limited physical activity include motor deficits, obesity, psychosocial disorders, behavioural problems and cardiovascular disease [21, 34, 39, 43, 48, 57]. Hence, being involved in physical activities has been considered a central health resource [18].

Physical fitness can be defined as the ability of a person to be physically active and includes endurance, strength, agility, coordination and flexibility. Physical fitness in children and adolescents has received international attention since the 1950s [1, 12, 16, 45, 46]. Beunen et al. and Telama conducted large-scale studies on fitness in Europe-

A. Woll (✉)
Universität Konstanz,
Sportwissenschaft | (Fach 30),
78457 Konstanz, Germany
e-mail: alexander.woll@uni-konstanz.de

B.-M. Kurth
Robert Koch-Institut,
Berlin, Germany

E. Oppen · K. Bös
Karlsruhe Institute of Technology,
Karlsruhe, Germany

A. Worth
PH Schwäbisch-Gmünd,
Schwäbisch-Gmünd, Germany

DEFENDANT'S
EXHIBIT

24

an children and adolescents [5, 47], yet only a few studies have followed participants until today [36]. A large percentage of children and adolescents present with poor physical fitness including coordination, strength, endurance and flexibility [55] and should thus receive special attention and intervention. In addition, there is some evidence that physical fitness including coordination and endurance in children and adolescents has declined [8, 25]. However, to date, conclusions regarding actual physical fitness level in children and adolescence cannot be drawn because of the lack of comprehensive and representative data.

Many different tools have been used for the assessment and comparison of physical fitness. However, even large-scale tools such as the Eurofit test battery do not assess all aspects of physical fitness [14]. Further, while physical fitness in children and adolescents has been compared between some countries [15, 31, 33, 44, 52], to date an internationally accepted, reliable standard tool for the assessment of physical fitness in children and adolescents is not available. Consequently, comparisons between existing data sets are challenging. In addition, levels or criteria of physical fitness in children and adolescents are needed to establish specific recommendation guidelines for interventions aimed at improving physical fitness.

Physical activities are the main leisure activities in children and adolescents [20]. Physical activity behaviour depends on socio-demographic parameters including age, gender, social status, education, migration background and urban versus rural environment [6, 40, 54], and these aspects are important in the context of physical activity and how it relates to health. Although more than 80% of German children and adolescents are members of a sports club at some point during their childhood and adolescence [20], the extent of their daily physical activities has decreased. A reduction of physical activity in children and adolescents appears to occur at a younger age. In addition, children not only become active members in sports clubs at a younger age but also end their active sports involvement at a younger age [28]. An important fact for the problem of health-oriented physical activity is that an estimated 75% of European adolescents between the ages of 11–15 years do not meet the guidelines for healthy activity behaviour recommended by the WHO [56]. However, previous studies on physical activity in German children and adolescents did not include all social contexts (school, club and non-organized) [9], did not differentiate between different types of physical activity [13] or are not representative of the entire population [24].

In general, representative prevalence data on physical activity in German children and adolescents are lacking, and representative longitudinal studies relating activity behaviour in German children and adolescents to their activity behaviour during adulthood and to their later health development are not available. Therefore, the purpose of

the ‘Motorik-Modul’ (MoMo) was to establish prevalence measurements on physical fitness and physical activity in German children and adolescents and to identify differences between age groups and genders.

Materials and methods

The ‘Motorik-Modul’ is supported by the German Ministry for Family, Seniors, Women and Youth and part of the German Child and Adolescent Health Survey (KiGGS) [27] conducted by the Robert-Koch-Institute in Berlin, Germany. Both studies were approved by the institutional review board and have been performed according to the Declaration of Helsinki. From 2003 to 2006, data on physical fitness and physical activity of 4,529 children and adolescents between the ages of 4–17 years from 167 cities across all states of the German Federation were collected. All participants or their guardians gave written informed consent prior to study participation. Participants were recruited from the KiGGS population, allowing for future analyses not only of data collected as part of the ‘Motorik-Modul’ but also their relationship to biomedical and psychosocial data collected as part of KiGGS (Table 1). To obtain a study population that is representative for children and adolescents with primary residence in Germany [27], the Robert-Koch-Institute (Berlin) and the Centre for Surveys, Methods and Analyses (ZUMA, Mannheim) used a stratified multi-stage probability sample with three evaluation levels where level 3 applied to the ‘Motorik-Modul’. To improve the representativeness of the study results, deviations from the sample from the population

Table 1 Biomedical and psychosocial data collected as part of KiGGS

Parameters
Information on physical status
Acute and chronic conditions
Disabilities
Health risks and accidents
Psychological health and behavioural problems
Subjective health and quality of life
Social contacts
Social network and support system
Personal resources such as protection mechanisms for a healthy psychological development
Nutrition, eating disorders and adiposity
Thyroid health and iodine intake
Health behaviour and leisure activities
Medication consumption and immunization status
Utilization of medical services

structure regarding age, gender, region and country of citizenship were corrected by weighing the data.

Teams of three testers from a total of 23 testers collected data at each location during two to three testing days, and each subject was involved for approximately 1 h. Testers were trained in data collection of motor tests, questionnaire administration and logistics after completing a first-aid or a defibrillation course. Each team consisted of at least one person certified in first-aid, one person certified in defibrillation and one female tester. The bike endurance test in girls between 6 and 10 years of age was always conducted by a female tester. Testers received continuing training to resolve possible questions arising during field testing. All testers received a tester identification code and were responsible for data entry to facilitate correct data collection and entry and to assess potential tester effects.

Sociodemographic parameters

The sociodemographic parameters included in this study were gender, age, BMI, social status, migration background, residential region (former German Democratic Republic (GDR) versus former Federal Republic of Germany) and residential environment (urban versus rural). All sociodemographic parameters were assessed in the umbrella study KiGGS and are important for the differential analysis of the physical fitness, physical activity and health status of children and adolescents. Differences between groups may reveal problem groups such as, for instance, children with activity deficits, fitness deficits or relevant environmental variables and hence provide prevalence information for the design of intervention or prevention programmes.

Table 2 Structure of physical fitness tests relating motor skills to motor abilities

Task structure		Motor abilities				Passive systems for transfer of energy
		Endurance	Strength	Agility	Coordination	Flexibility
		aerobic	strength endurance/ power	reaction time	coordination under time constraint/ coordination with precision	flexibility
Gross motor skills						
Locomotion	walking				backward balancing(BAL)	
	jumping		standing long jump (SW) force plate (KMP)		jumping side-to-side (SHH)	
Gross motor limb movements	upper extremities		push-ups (LS)			forward bend (RB)
	trunk					
	lower extremities	bike endurance test (RAD)				
Posture	whole body				single leg stance (EINB)	
Fine motor skills						
Fine motor limb movements	hand			reaction test (REAK)	MLS-tracing lines (LIN) MLS-sorting pens (STI)	

Physical fitness tests

While different tools have been used to assess physical fitness in children and adults [15, 31, 33, 44, 52], to date a generally accepted standard tool has not been established. The criterion for the selection of physical fitness tests was the ability to assess as many aspects of physical fitness as possible. The general physical fitness test structure relating motor skills to motor abilities is described in Table 2. Single leg stance was used for assessing gross motor coordination during static precision tasks [42]. The subjects were asked to stand on their dominant leg for 1 min with their eyes open, and the number of floor contacts with the contralateral limb was recorded. The backward balancing was based on a body coordination test [41] and allowed the assessment of gross motor coordination during dynamic precision tasks. The subjects were asked to balance backwards on 6-, 4.5- and 3-cm-wide beams, respectively, with two trials per beam. The numbers of steps on each beam were added. The test was terminated if one foot touched the ground. Fine motor coordination during precision tasks was assessed using the tracing lines and sorting pens tests. The subjects were asked to trace a groove on a board using a pen without touching the rims of the groove. The average time of tracing without touching the rims was calculated. The subjects were then asked to move 25 pens from a pen holder into wholes at the edge of a board as quickly as possible. Reaction time was assessed using a computer-aided reaction test. The subjects were asked to react as quickly as possible to 14 colour changes of a traffic light, and the seven fastest times were averaged. Forward bends [26] were used for the assessment of trunk flexibility and the flexibility of the sciatic crural muscle group. The lowest point reached by the fingertips while standing on a box with legs extended was recorded. Positive values indicated good flexibility while negative values indicated poor flexibility. Standing long jumps [23] and counter-movement jumps on a force plate that was developed by our group were used to assess leg power. Dynamic strength endurance of the upper extremities was assessed using the push-up test [4]. The subjects were asked to do as many push-ups as possible within 40 s. The jumping side to side test [41] was used to assess gross motor coordination under time constraint. The subjects were asked to perform as many jumps from side to side as possible during two 15-s intervals, and the numbers for the two intervals were averaged. Aerobic endurance was tested using the bike endurance test with test parameter PWC170 [38]. The initial load was adjusted to 0.5 W/kg body mass. Subsequently, the load was increased every 2 min by 0.5 W/kg increments. The test was abandoned when the predetermined threshold heart rate was reached for at least 15 s (190 and 180 heartbeats/min for children aged 6–10 years and children and adolescents aged 11–17 years, respectively).

In a pilot study on 138 female and male subjects between the ages of 4–17 years, the quality of the physical fitness tests was assessed using the test–retest method [32]. The correlation coefficient for results for different testers was 0.98 or 0.99 for all tests. The percent difference between results for different testers was less than one percent point. The overall reliability for the standardized overall values was $r=0.971$ ($p<0.001$). Overall, the selection of test tasks was sufficiently reliable and valid for the assessment of physical fitness in children and adolescents 4–17 years of age.

Physical activity assessment

A questionnaire consisting of 35 items was developed for the assessment of physical activity. The main categories of the ‘Motorik-Modul’ questionnaire are given in Table 3. The first two items assessed overall physical activity for at least 60 min per day during the preceding 7 days at moderate to high intensity and facilitated comparisons with internationally developed activity standards (e.g. 60 min per day). The remaining 33 items covered the following activity categories: daily activities, physical activity classes at school, club sports and physical activities outside of clubs. Information on duration and frequency was collected for each activity category. In addition, information on perceived intensity (1—no, 2—mild, and 3—high perspiration and shortness of breath, respectively) was collected for physical activities (at school, clubs or outside of clubs). For all categories except physical activity at school, information on the type of physical activity was collected. While the activity type for daily activities was categorized into playing outdoors, garden and farming activities and walking as mode of transportation, the subjects could list any type of activity for club and leisure activities. To obtain detailed information, answers were given as duration (minutes per day) and frequency (days per week). Because the subjects do not participate in all activities during the entire year (e.g. skiing), separate items assessed the months of each activity (e.g. January–December) with the goal of a precise assessment of activity behaviour in terms of habitual activity. Finally, a range of predicting parameters were assessed, including accessibility of sports facilities, physical activity behaviour of next of kin, health-related expectations, physical expectations, socio-emotional expectations and motive for being physically active [19]. The reliability and validity of the final instrument was tested in a pilot study [32] using a test–retest. The reliability of the instrument was $r=0.97$, and no significant differences in mean values between test and retest values were detected. The validity of the instrument was tested by comparing activity monitor and heart rate monitor data to values obtained using the instrument. The agreement between information obtained using these systems was

Table 3 Main categories of the ‘Motorik-Modul’ questionnaire for the assessment of physical activity [37]

Content	Question	Source
General physical activity	On how many days during the last week were you physically active for at least 60 min/day? On how many days of a normal week are you physically active for at least 60 min per day?	Prochaska, Sallis & Long [35]
Physical activity at school	On how many days per week do you have physical education in school? How many lessons (45-min lessons) per week does this usually amount to? How strenuous are these lessons usually for you? Are you a member of a school sports team? If yes, which sports team are these? How many lessons of sports team practice (45-min lessons) per week does this usually amount to?	Ulmer [52] Bös et al. [7] Bös et al. [7] Ulmer [52]
Daily physical activities	How do you usually get to school? How often do you usually play outside during a typical week? Do you help out in the garden or with farm work (e.g. picking fruit, etc.)? How far do you usually walk on a daily basis?	no reference Bös et al. [7] Ulmer [52]
Physical activity in a sports club	Are you a member of a sports club? Which sport(s) do you actively do in a club? How many times per week do you participate in each sport? How long does a training session last? During which months do you participate in each sport? How strenuous are these practises typically? Do you participate in competitions?	Ulmer [52] Bös et al. [7] Ulmer [52], Bös et al. [7]
Physical leisure activities outside of sports clubs	Do you participate in any other sports outside of sports clubs? Which sport(s) do you actively do outside of sports clubs? How many times per week do you participate in each sport (outside of sports clubs)? How many minutes per week do these activities usually amount to? During which months do you participate in these sports? How strenuous are these activities typically?	Kurz, Sack & Brinkhoff [28] Ulmer [52] Fuchs [17]
Availability of sports facilities	The closest sports facility is far from my home When I would like to do sports, I am lacking the facilities and equipment	Fuchs [17]
Interest in physical activity at school	How interested are you in physical activity at school?	Bös et al. [7]
Physical activity behaviour of next of kin	Is your dad regularly physically active? Is your mom regularly physically active? Are your siblings regularly physically active? How many of your friends are regularly physically active?	Fuchs [17]
Health-related expectations	When I am physically active regularly, ...I won't get sick as easily ...I can easily get injured	Fuchs [17]
Physical expectations	...my endurance will improve ...my joints will stay flexible ...I will have a nice body and look nice	
Socio-emotional expectations	...I spend time with others ...I get rid of my energy ...I won't be bored ...I will not think about my problems	
Motivation for being physically active	I do sports.... ...to have fun ...to do something with other people ...to do something about my health ...to relax ...to stay in shape ...to get rid of stress ...to compete with others ...to stay fit ...to get better ...other	Ulmer [52]

Table 4 Study population by age and gender (4–17 years)

Age [years]	Boys	Girls	Total
4	153	146	299
5	155	146	301
6	158	149	307
7	161	153	313
8	158	151	309
9	152	147	299
10	156	142	298
11	157	154	311
12	158	154	312
13	174	160	334
14	180	179	359
15	186	176	362
16	194	178	372
17	172	178	350
Total (4–17 years)	2,314	2,215	4,529
Percent (total)	51.1%	48.9%	100%

small for low-intensity activities ($r=0.26$) to moderate for high-intensity activities ($r=0.51$) and comparable to results reported in the literature [7, 17, 35].

Statistical analyses

All statistical tests were conducted using SPSS Version 16.0 (SPSS Inc., Chicago, IL, USA). Chi-square tests were used to detect differences in nominal and ordinal variables between groups. Univariate analyses of variance were used to detect differences in physical fitness and physical activity data between subject groups (age, gender). Relevance of a mean difference was determined by a difference of at least 0.5 SD or 5%. Data were compared with criterion-related guidelines.

Results

Sociodemographic parameters

The age and gender distribution details of the study population are shown in Table 4. Of all subjects, 18.4% of subjects lived in rural areas, 27.9% in small towns, 29.4% in small cities and 24% in major cities, respectively. Further, 14.3% of participants had a migration background, and 13.6% lived in states that formerly belonged to the GDR. Finally, 24.6% of subjects belong to the lower class, 46.8% to the middle class and 27.2% to the upper class. Overall, the study sample was representative of children and adolescents residing in Germany with regards to age, gender, region and country of citizenship according to data

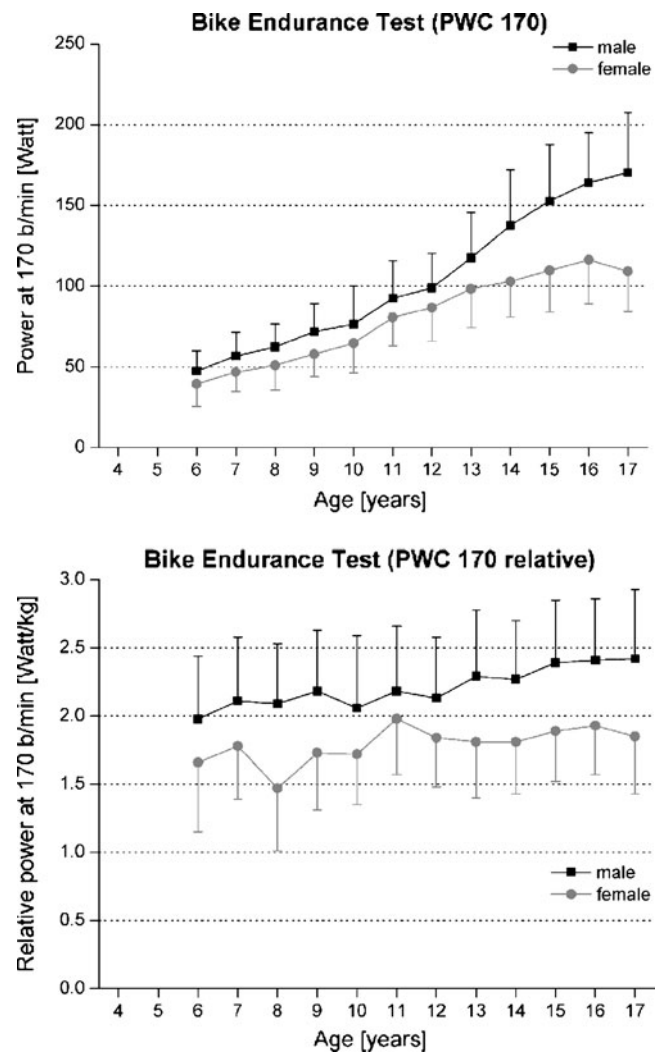


Fig. 1 Endurance tests. Mean (1 SD) power at 170 b/min during the bike endurance test by age and gender ($N=3,410$). *Top graph*—absolute power. *Bottom graph*—relative power (normalized to body mass)

provided by the Centre for Surveys, Methods and Analyses (ZUMA, Mannheim).

Physical fitness

Eighty-seven percent of the 3,528 children and adults tested reached the threshold heart rate. Thirteen percent terminated the test prematurely because of subjective exhaustion or lack of motivation. For 208 (5.7%) of subjects, PWC170 could not be determined. The mean values of absolute and relative endurance performance were greater in older subjects of both genders ($P<0.001$; Fig. 1). In all age groups, male subjects performed better than female subjects ($P<0.001$). A significant interaction between the factors age and gender was identified ($P<0.001$) where the differences were greater in boys over the age of 10 years while the differences were linear in girls of all ages. On average,

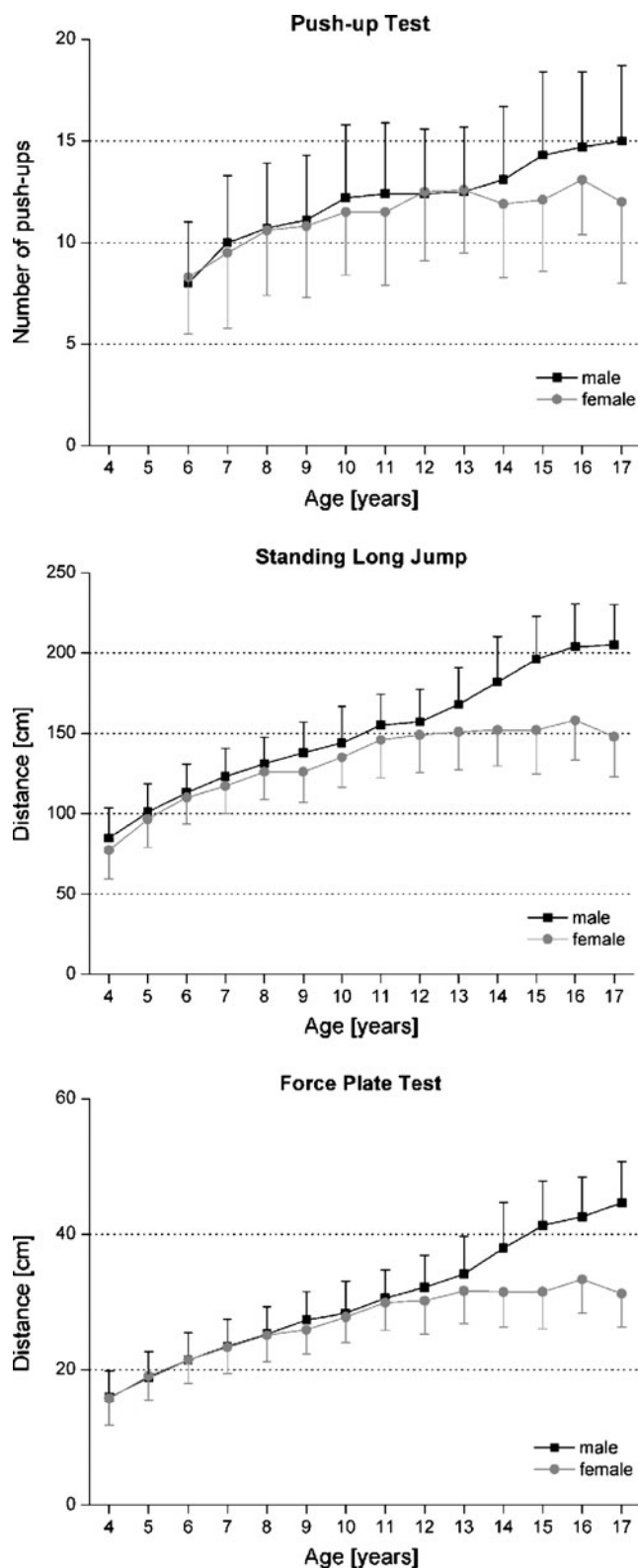


Fig. 2 Strength tests: *top graph*—mean (1 SD) number of push-ups performed during 40 s by age and gender ($N=3,887$). *Centre graph*—mean (1 SD) distance achieved during the standing long jump test by age and gender ($N=4,504$). *Bottom graph*—mean (1 SD) distance achieved during the force plate test with bilateral push-off, arms akimbo and no arm swing

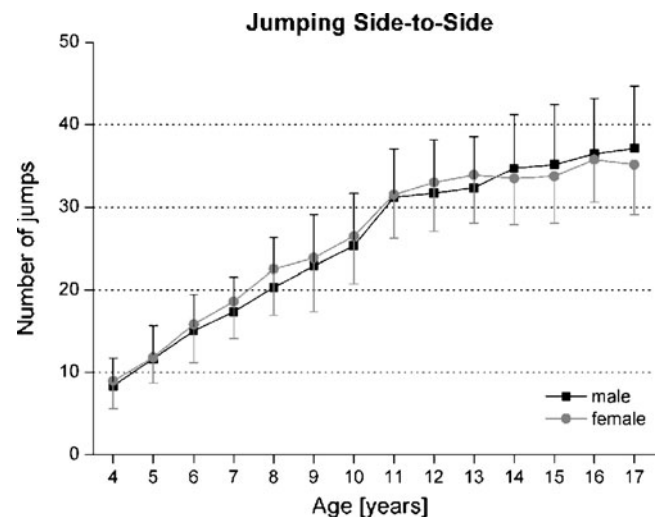


Fig. 3 Test for coordination under time constraint. Mean (1 SD) number of side-to-side jumps performed during two 15-s intervals by age and gender ($N=4,471$)

the difference in performance is 11 W/year for boys and 6 W/year for girls.

All strength tests showed that strength is higher the older the subjects are ($P<0.001$; Fig. 2). While the difference in strength between age groups was constant for boys, the difference in strength between age groups for girls above the age of 12 years was negligible, causing a greater mean strength in boys aged 14–17 years than in girls in the same age groups ($P<0.001$; Fig. 2). Only in the standing long jump test did boys aged 4–13 years perform better than girls in the same age groups ($P<0.001$; Fig. 2).

The number of jumps in the jumping side to side test was higher in older subjects ($P<0.001$; Fig. 3). While below the age of 10 years the difference in the number of jumps between age groups was three jumps per year, this difference was only one jump/year in subjects aged 11–17 years. Results for boys and girls were similar in all age groups.

The number of ground contacts for the single leg stance test was lower and the number of points for the backward balancing was higher in older subjects ($P<0.001$ for both; Fig. 4). Both tests revealed better test performance for girls than for boys in all age groups ($P<0.001$ for both). The difference in test results between age groups was similar for both genders. While the difference in test performance between age groups was close to constant for children 11 years and younger, these differences were negligible for age groups 12 years and older. Both tests differentiate well only in younger children.

Test performance in all tests for fine motor coordination was higher in older subjects ($P<0.001$; Fig. 5). While boys and girls showed similar test performance for the reaction test, girls performed better than boys in the tracing lines and

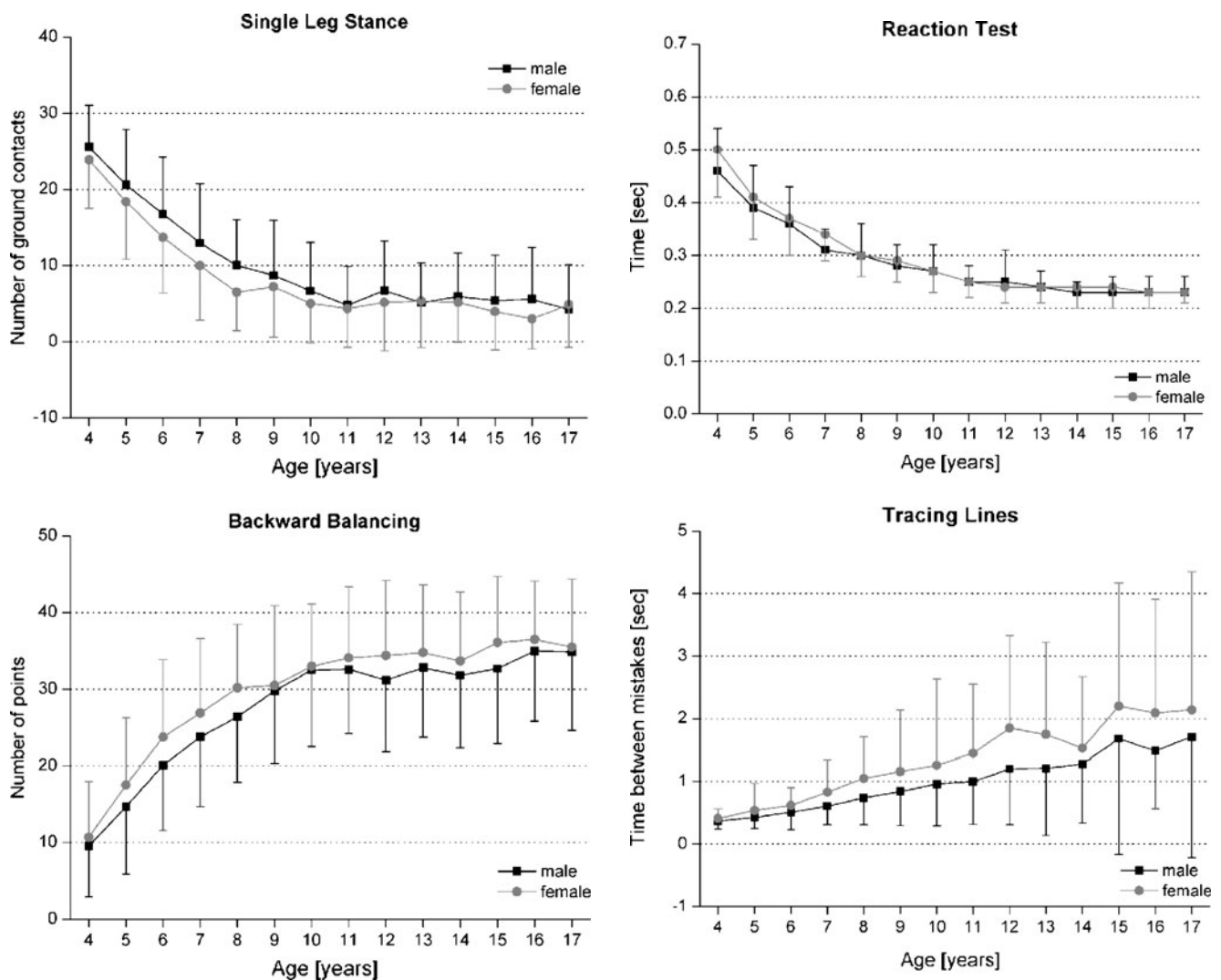


Fig. 4 Tests for coordination during precision tasks. *Top graph*—mean (1 SD) number of ground contacts during a 60-s single leg stance trial on a T-shaped beam by age and gender ($N=4,459$). *Bottom graph*—mean (1 SD) number of points achieved during six backward balancing trials on three different width beams by age and gender ($N=4,502$)

sorting pens tests ($P<0.001$ for both). The difference in test performance between age groups was similar for both genders for the reaction test and the sorting pens test and was negative in older subjects. Differences in test performance for the tracing lines test between age groups were almost constant for both genders, but the difference between age groups was greater in girls than in boys ($P=0.006$).

Values for the forward bend test were consistent between all age groups (Fig. 6). In all age groups, girls performed better than boys ($P<0.001$).

Physical activity

On average, children and adolescents participated in 2.2 ± 0.9 h per week on 1.6 ± 0.9 days per week of physical activity

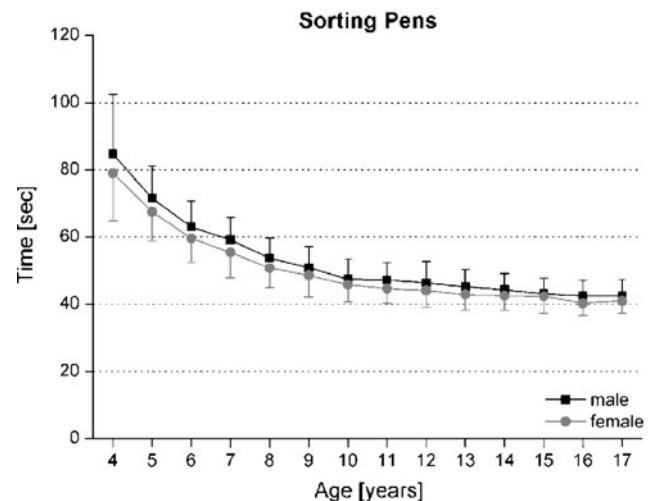


Fig. 5 Tests for fine motor coordination. *Top graph*—mean (1 SD) reaction time to colour changes of a traffic light by age and gender ($N=4,467$). *Centre graph*—mean (1 SD) time between mistakes while tracing a groove using a pen by age and gender ($N=4,406$). *Bottom graph*—mean (1 SD) time needed to move 25 pens by age and gender ($N=4,401$)

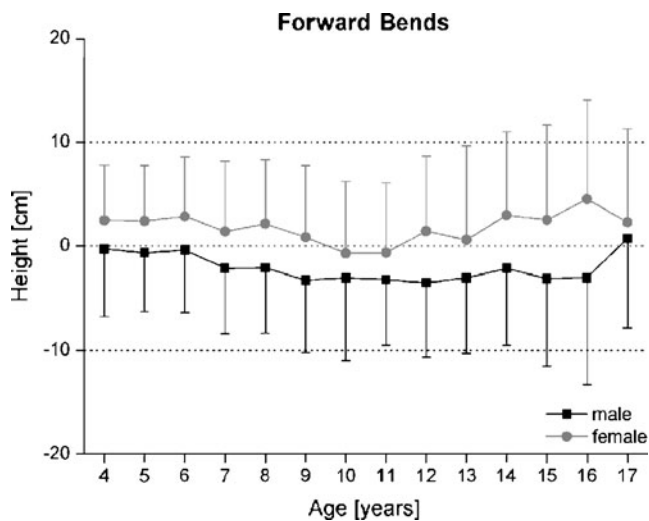


Fig. 6 Flexibility test: mean (1 SD) of height reached during the forward bending test by age and gender ($N=4,464$)

at kindergarten and school with a significant age effect ($P<0.001$ for both; Table 5). In addition, 10.8% of boys and 9.0% of girls participated in elective physical activities at school (Table 5). Younger children participated in outdoor physical activities on more days per week than older children and adolescents ($P<0.001$) did, and boys were more frequently active outside than girls of the same age ($P<0.001$; Table 5). On average, approximately 60% of boys and girls participated in non-club physical activities with an average of almost 2.5 h per week (Table 5). While 63% of boys and 52% of girls are members in a sports club, 16.5% of boys and 21.0% of girls have opted out of their previous club membership, a phenomenon that increases with increasing age ($P<0.001$; Table 5). The time spent on club sports and the percentage of children participating in competitions was greater in older subjects and greater for boys than for girls ($P<0.001$ for both; Table 5). Nevertheless, only 17.4% of girls and 13.1% of boys fulfil the requirement put forward by the WHO in 2002 (Tables 5 and 6).

Girls perceived physical activity at kindergarten or school, non-club and club physical activities as more intense than boys ($P<0.001$ for all; Figs. 7, 8 and 9). The intensity of physical activity at kindergarten and school, non-club and club physical activity was smaller in older subjects for both genders ($P<0.001$ for all; Figs. 7, 8 and 9). While the percentage of moderate-intensity activities remained relatively constant across all age groups for physical education at kindergarten and school and for non-club physical activities (Figs. 7 and 8), this percentage was smaller in older subjects for club sports with an associated greater amount of low-intensity activities for this type of physical activity (Fig. 9).

Table 5 Physical activity in children and adolescents by age group. Mean (1 SD) values are presented. P -values for age effect are given

Category	Parameter	Count, N	4–5 years old	6–10 years old	11–13 years old	14–17 years old	Overall	P -value
Activities at school or kindergarten	Number of hours/week spent on physical activity	4,426	1.5 (1.2)	2.4 (0.8)	2.5 (0.8)	2.1 (0.9)	2.2 (0.9)	<0.001
	Number of days/week of physical activity	4,071	1.3 (1.0)	2.0 (0.8)	1.7 (0.8)	1.3 (0.8)	1.6 (0.9)	<0.001
Daily activities	Participation in elective sports classes (boys/girls) [%]	3,879	N/A	11.0/7.1	13.0/10.3	10.0/9.6	10.8/9.0	n.s.
	Number of days/week with outdoor activities (boy/girls)	4,417	6.2/5.8	5.5/5.3	4.5/4.2	3.3/2.7	4.6/4.2	<0.001 (for age, gender)
Activities outside sports clubs	Participation in non-club sports (boys/girls) [%]	4,394	50/49	60/58	64/66	67/61	61.7/59.5	<0.001
	Number of minutes per week for non-club sports (boys/girls)	4,219	78/59	117/96	178/109	208/127	156/106	<0.001 (for age, gender)
Activities in sports clubs	Membership in sports clubs (boys/girls) [%]	4,502	50.0/54.5	71.4/57.3	68.8/52.8	55.8/21.7	63/52	<0.001 (for age, gender)
	Opt out of sports clubs (boys/girls) [%]	842	5.7/5.9	8.3/12.4	16.6/25.2	29.7/33.4	16.5/21.0	<0.001 (for age, gender)
	Number of minutes per week for club sports (boys/girls)	2,451	90/82	150/123	219/183	256/219	190/157	<0.001 (for age, gender)
	Participation in competitions (boys/girls) [%]	2,247	22.7/9.9	65.5/36.3	74.2/67.6	79.6/68.5	67.0/48.0	<0.001 (for age, gender)
Fulfillment of WHO activity guidelines	Boys/girls [%]	3,943	35.4/28.4	24.2/17.9	9.4/8.3	8.1/5.0	17.4/13.1	<0.001 (for age, gender)

Table 6 Comparison between results of the ‘Motorik-Modul’ and WHO with regards to activity guidelines

Guidelines	Gender	Age group								
		11 years			13 years			15 years		
		MoMo 2006	WHO 2002	WHO 2008	MoMo 2006	WHO 2002	WHO 2008	MoMo 2006	WHO 2002	WHO 2008
Number of days being active for at least 1 h	M	3.9	4		3.9	3.8		3.4	3.8	
	F	3.4	3.5		2.9	3.3		3.0	3.2	
Being active for 1 h on 5 days per week [%]	M	34.6	35.3		31.3	30.8		25.1	29.4	
	F	26.0	23.5		16.8	18.7		18.2	17.9	
Being active for 1 h on 7 days per week [%]	M	16.7		25.0	20.4		19.0	12.6		16.0
	F	15.3		20.0	9.6		13.0	12.7		10.0

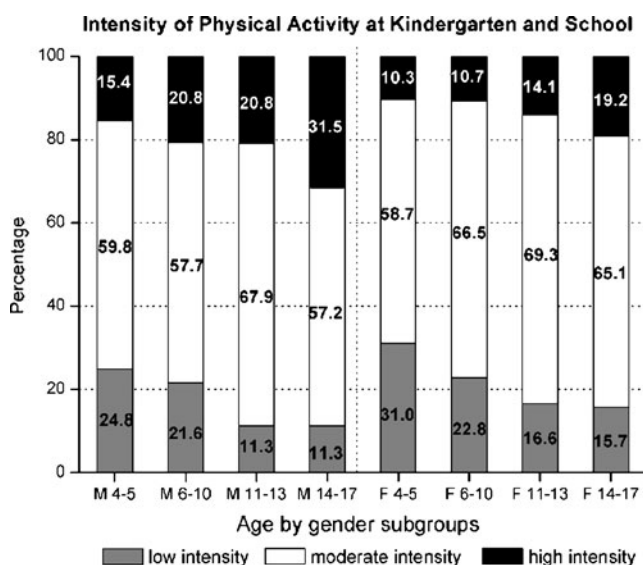
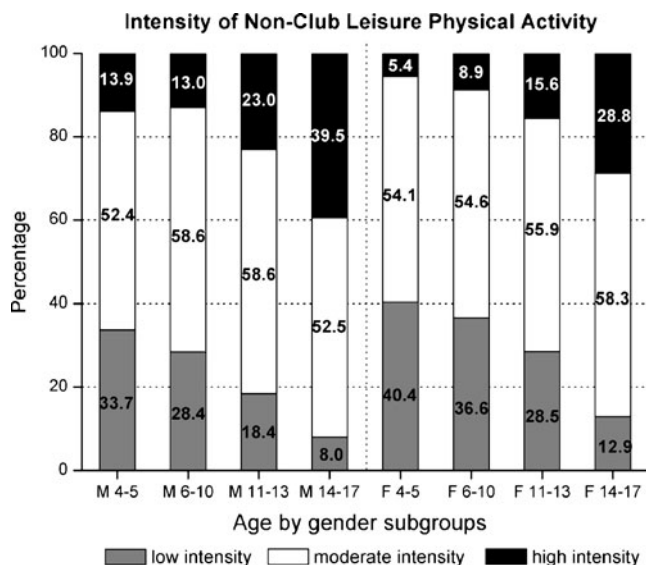
Soccer and swimming were the only sports that were listed as the favourite club and non-club physical activities by boys of all ages, and swimming was the only sports listed as their favourite sports by girls of all ages (Tables 7 and 8). While 4- to 5-year-olds by far favoured biking as non-club physical activity and gymnastics as club sports, the distribution among different sports became more even between the top five sports with increasing age, especially for girls (Tables 7 and 8).

More than 85% of children and adolescents walked 1 km or more per day (Fig. 10). Only 9.8% of children and adolescents walked more than 6 km per day. Walking distance per day was greater in older subjects ($P<0.001$) but did not differ between boys and girls.

Discussion

The purpose of this study was to establish prevalence measurements on physical fitness and physical activity in German children and adolescents and to identify differences between age groups and genders. A large data set on physical fitness test results and physical activity information was presented. Common age and gender-dependent patterns were observed for many parameters and tests describing physical fitness and for physical activities in different settings.

For parameters describing physical fitness, age was the dominant influencing variable. Only for forward bends was the effect of gender stronger than the effect of age [3]. In general, endurance and strength increased from childhood to puberty as reported in the literature [22, 30]. While endurance and strength continues to increase in boys older

**Fig. 7** Intensity of physical activity at kindergarten and school by age and gender (M, boys: $N=2,206$; F, girls: $N=2,111$)**Fig. 8** Intensity of non-club physical activities by age and gender (M, boys: $N=1,372$; F, girls: $N=1,279$)

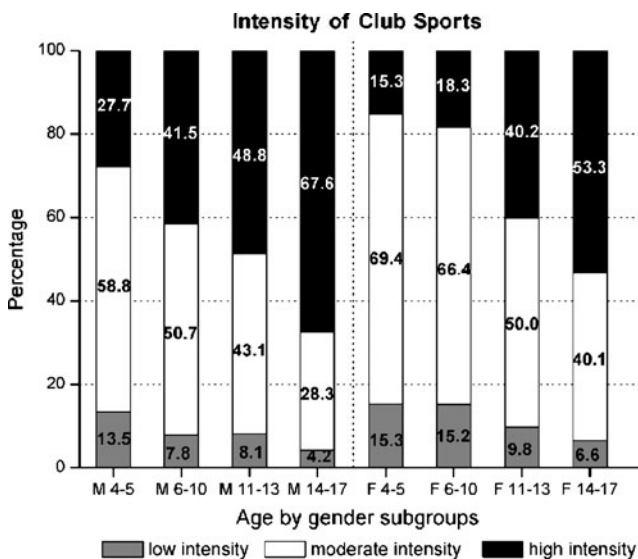


Fig. 9 Intensity of club sports by age and gender (M, boys: $N=1,446$; F, girls: $N=1,113$)

than 12–13 years, levels remain constant for girls in this age range [30]. Only aerobic endurance continued to improve throughout puberty in girls; however, these improvements occurred at a slower rate for girls than for boys, resulting in a gender-specific pattern.

While girls excelled in gross motor coordination compared to boys at all ages, no gender difference was found for gross motor coordination under time constraints. In general, gross motor skills improved from childhood to puberty and reached a plateau during adolescence. In contrast to strength and endurance, values started to be constant at age 11–12 years. Similarly, reaction time and fine motor coordination under time constraint was only slightly better in older subjects. However, fine motor coordination during precision tasks continued to increase throughout adolescence [53], describing age difference due to neuromotor development [29].

Activity behaviour in children and adolescence is of special interest with the increasing influence of technology in our daily lives and the importance of being physically active for our long-term health [2]. While the education minister of the German federation recommends at least 3 h of physical education at schools, these recommendations are not always met. In fact, similar to the results of the SPRINT study [9], the results of this study revealed that only 2.2 h per week of physical education are taught at German schools. Similar to previous reports [28, 52], less than 10% of children and adolescents participated in elective physical activity classes and because of their participation in elective physical activity classes these children typically exceed the recommended hours of physical education at school. Younger children and boys

Table 7 Order of favourite sports outside of sports clubs by age and gender in percent

Sports	Boys [%]	Girls [%]
4–5 years (M, $N=145$; F, $N=138$)		
Biking	57.7	52.3
Soccer	30.9	
Swimming	22.6	33.1
Inline skating	7.7	11.6
Dancing		10.7
Skiing	6.5	7.8
6–10 years (M, $N=455$; F, $N=420$)		
Biking	63.6	61.4
Soccer	31.7	
Swimming	30.6	32.1
Inline skating	23.4	36.5
Horseback riding		6.9
Skiing	10.2	7.9
11–13 years (M, $N=303$; F, $N=300$)		
Soccer	43.4	13.4
Biking	38.6	42.7
Swimming	18.5	17.6
Inline skating	13.0	29.7
Jogging, running		13.6
Basketball	12.5	
14–17 years (M, $N=477$; F, $N=427$)		
Soccer	34.1	
Biking	28.8	31.4
Basketball	19.2	
Jogging, running	13.1	26.2
Inline skating		20.3
Dancing		9.5
Swimming	10.4	19.2

had a greater desire of being physically active as previously reported [49, 50], and all children and adolescents performed these activities predominantly outdoors and as leisure activities.

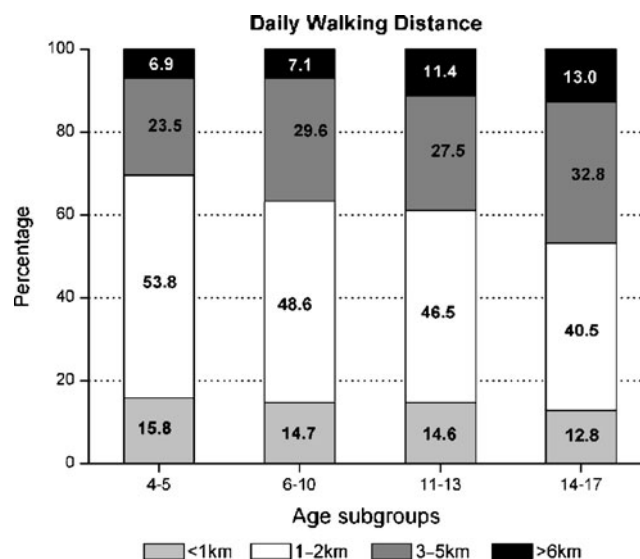
In Germany, there are 90,000 registered sports clubs, and 56,500 of these clubs offer physical activity programmes for children under the age of 7 years and 84,000 offer physical activity programmes for children and adolescents aged 7–17 years [10]. Membership fees are affordable, sports clubs exist even in rural areas and programmes in a variety of sports are offered. The attractiveness of these programmes is reflected in the large percentage of children and adolescents who are members in sports clubs. On average, children and adolescents spend almost 3 h per week participating in clubs sports. However, many clubs appear to have difficulty in being attractive for adolescence, and 30% of adolescents opt out of their club membership. These results emphasize the

Table 8 Order of favourite club sports by age and gender in percent

Sports	Boys [%]	Girls [%]
4–5 years (M, N=150; F, N=157)		
Gymnastics	46.6	55.6
Soccer	38.3	
Dancing	29.7	
Swimming	11.4	15.5
Handball	5.3	2.0
Tennis	3.7	2.0
Track and field		2.9
6–10 years (M, N=558; F, N=422)		
Soccer	52.0	
Swimming	17.9	19.9
Dancing		26.0
Gymnastics	10.1	31.5
Track and field	9.6	8.0
Tennis		
Handball	7.7	
11–13 years (M, N=333; F, N=247)		
Soccer	47.1	
Gymnastics		15.2
Table tennis	11.7	
Horseback riding		13.5
Swimming	11.2	18.5
Handball	6.1	
Dancing		12.8
Volleyball	4.0	9.3
14–17 years (M, N=407; F, N=317)		
Soccer	42.9	8.9
Dancing		29.9
Tennis	12.7	
Horseback riding	13.7	
Table tennis	7.0	
Swimming	6.7	8.2
Volleyball		10.2
Handball	6.1	

importance of physical education at schools especially for children and adolescents aged 11–17 years. In contrast, those who maintained their membership spent up to more than 4 h per week on training. This development is likely related to the increasing time required for preparation for competitions as three out of four adolescents who were members in a sports club also participated in competitions.

Physical activity performed at a club usually occurred with a higher intensity than physical activity at kindergarten and school and during leisure time, and intensity increased with increasing age. However, in this study, subjective measures of intensity were used. Hence, it is unclear if the required intensity differed between settings or if children

**Fig. 10** Average walking distance per day for both genders by age (N=3,878)

and adolescents perceived the intensity differently in different settings. Nevertheless, it appears that those children and adolescents who actively participated in club sports did their sports with moderate to high intensity. This result is in agreement with previous reports of greater intensity of after-school physical activity than that during physical activity classes in school in American children [51]. The fact that the type of club sports favoured by children and adolescents remains relatively consistent throughout childhood and adolescence suggest an association of age-related increasing intensity with competition involvement rather than with type of sports.

Only 15.3% percent of children and adolescents fulfilled the recommended activity level of 60 min of physical activity with moderate intensity on 7 days per week, and the younger the age group the more children fulfilled these requirements. These results suggest that, with increasing institutionalization of the child's daily routines, longer school days and increasing time requirements for homework, less and less time is available for daily physical activities. Simultaneously, physical activity shifts from the daily activity setting to organized sports at sports clubs on a few days per week with a change in frequency and quality of physical activity. These developmental processes explain the apparent paradox that children are more physically active while less frequently fulfilling activity requirements with increasing age. Hence, the question arises if the activity guidelines are appropriate measures for amount of activity and if a 1-h daily activity at moderate intensity is better with regards to health than an activity that is performed less frequently but at a higher intensity. Nevertheless, the results for German children and adoles-

cents are in general agreement with those reported by the WHO and by other studies [11, 56].

In summary, this large-scale cross-sectional study provides prevalence data on physical fitness and physical activity in children and adolescents. Age and gender differences were identified for many parameters describing physical fitness and physical activity. The relevance of these differences will become apparent when the current data set is put into relationship with other aspects such as socioeconomic factors and health information. In addition, longitudinal data building upon the current data set will provide critical information on the predictive value of physical fitness and physical activity behaviour for an individual's long-term health.

Acknowledgements The authors thank PD Dr. habil. Annegret Mündermann (Ph.D.) for her writing assistance. This study was funded by the German Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research).

Conflict of interest The authors do not have a financial relationship with the organization that sponsored the research.

References

1. American Alliance for Health PEaRA (eds) (1965) Youth fitness test. Manual. American Alliance for Health, Physical Education and Recreation, Washington, DC
2. Andersen LB, Harro M, Sardinha LB et al (2006) Physical activity and clustered cardiovascular risk in children: a cross-sectional study (The European Youth Heart Study). *Lancet* 368:299–304
3. Araujo CG (2008) Flexibility assessment: normative values for flexitest from 5 to 91 years of age. *Arq Bras Cardiol* 90:257–263
4. Beck J, Bös K (1995) Normwerte motorischer Leistungsfähigkeit. Sport und Buch Strauss, Köln
5. Beunen G (1982) The Leuven Growth Study of Flemish Girls: study design and data quality control. In: Simons J, Renson R (eds) Evaluation of motor fitness. Report of the European research seminar on the evaluation of motor fitness. Institut of Physical Education, Leuven, pp 59–76
6. Borraccino A, Lemma P, Iannotti RJ et al (2009) Socioeconomic effects on meeting physical activity guidelines: comparisons among 32 countries. *Med Sci Sports Exerc* 41:749–756
7. Bös K, Heel J, Romahn N et al (2002) Untersuchung zur Motorik im Rahmen des Kinder- und Gesundheitssurveys [Examination of motor fitness within the scope of The Child and Adolescent Health Survey]. *Gesundheitswesen* 64(Suppl 1):S80–S87
8. Bös K, Opper E, Woll A (2002) Fitness in der Grundschule. Förderung von körperlich-sportlicher Aktivität, Haltung und Fitness zum Zwecke der Gesundheitsförderung und Unfallverhütung. Bundesarbeitsgemeinschaft für Haltungs- und Bewegungsförderung, Wiesbaden
9. Brettschneider W-D (2005) Die SPRINT-Studie Eine Untersuchung zur Situation des Schulsports in Deutschland. Meyer & Meyer, Aachen
10. Breuer C, Wicker P (2008) Sportvereine in Deutschland. Sportentwicklungsbericht 2007/2008. Analyse der Situation der Sportvereine in Deutschland. Deutsche Sporthochschule Köln, Cologne
11. Corbin CB, Pangrazi RP, Welk GJ (1998) Physical activity for children. A statement of guidelines. Council for Physical Education for Children (COPEC) of the National Association for Sport and Physical Education an association of the American Alliance for Health, Physical Education, Recreation and Dance
12. Davis KL, Roberts T, Smith RR et al (1994) North Carolina Children and Youth Fitness Study. *Am J Health Educ* 65:65–72
13. Deutsche Shell Holding GmbH (2002) Jugend 2002. 14. Shell Jugendstudie. Fischer, Frankfurt am Main
14. Eurofit (1993) Eurofit Tests of Physical Fitness. Strasbourg
15. Fares M (1982) Physical Fitness und sportmotorische Geschicklichkeit bei 11-14jährigen deutschen und ägyptischen Schülern. Ph.D. dissertation. Universität Heidelberg, Heidelberg
16. Fleishman EA (1964) The structure and measurement of physical fitness. Prentice Hall, New York
17. Fuchs R (1989) Sportliche Aktivitaet bei Jugendlichen. Entwicklungsverlauf und sozial-kognitive Determinanten [Sport activity in adolescents. Development and socio-cognitive determinants]. bps, Köln
18. Fuchs R, Göhner W, Seelig H (2007) Aufbau eines körperlich-aktiven Lebensstils. Theorie, Empirie und Praxis. Hogrefe, Göttingen
19. Fuchs R, Semmer N, Lang P et al (1987) Körperliche Aktivität von Jugendlichen. Validität der Messung, Prävalenz und Determinanten. In: Laaser U, Sassen G, Murza G et al (eds) Prävention und Gesundheitserziehung. Berlin, Springer, pp 245–256
20. Gogoll A, Kurz D, Menze-Sonneck A (2003) Sportengagements Jugendlicher in Westdeutschland. In: Schmidt W, Hartmann-Tews I, Brettschneider W-D (eds) Erster Deutscher Kinder- und Jugendsportbericht. Hofmann, Schorndorf, pp 145–165
21. Green LN, Williams K (1992) Differences in developmental movement patterns used by active versus sedentary middle-aged adults coming from a supine position to erect stance. *Phys Ther* 72:560–568
22. Holm I, Fredriksen P, Fosdahl M et al (2008) A normative sample of isotonic and isokinetic muscle strength measurements in children 7 to 12 years of age. *Acta Paediatr* 97:602–607
23. Kirsch A (1968) Standard fitness test. Lehrhilfen Leibeserzieh 17:133–140
24. Klaes L, Rommel A, Cosler D et al (2000) Bewegungsstatus von Kindern und Jugendlichen in Deutschland. Kurzfassung einer Untersuchung im Auftrag des Deutschen Sportbundes und der AOK- Die Gesundheitskasse. WIAD-Studie, Bonn
25. Knuth AG, Hallal PC (2009) Temporal trends in physical activity: a systematic review. *J Phys Act Health* 6:548–559
26. Kraus H, Hirschland RP (1954) Minimum muscular fitness tests in school children. *Res Q Exerc Sport* 25:178–188
27. Kurth BM, Kamtsiuris P, Holling H et al (2008) The challenge of comprehensively mapping children's health in a nation-wide health survey: design of the German KiGGS-Study. *BMC Public Health* 8:196
28. Kurz D, Sack H-G, Brinkhoff K-P (1996) Kindheit, Jugend und Sport in Nordrhein-Westfalen. Der Sportverein und seine Leistungen. Eigenverlag, Düsseldorf
29. Largo RH, Fischer JE, Rousson V (2003) Neuromotor development from kindergarten age to adolescence: developmental course and variability. *Swiss Med Wkly* 133:193–199
30. McMurray RG, Harrell JS, Bradley CB et al (2003) Gender and ethnic changes in physical work capacity from childhood through adolescence. *Res Q Exerc Sport* 74:143–152
31. Morrow JR (1992) Are American children and youth fit? Some international perspectives. *Res Q Exerc Sport* 63:449–452
32. Oberger J, Romahn N, Opper E et al (2006) Untersuchungen zur motorischen Leistungsfähigkeit und körperlich-sportlichen Aktivität im Rahmen des Kinder- und Jugendgesundheitsurveys des Robert Koch-Institutes Berlin [Study on motor fitness and

- physical activity as part of the German Health Interview and Examination Survey for Children and Adolescents of the Robert Koch-Institutes Berlin]. In: Wydra G, Winchenbach H, Schwarz M et al (eds) *Assessmentverfahren in Gesundheitssport und Bewegungstherapie*. Czwalina, Hamburg, pp 44–55
33. Papavassilou I (2000) Beurteilung der körperlichen Leistungsfähigkeit griechischer SchülerInnen mit Hilfe des IPPTP. Ph.D. dissertation. Universität Karlsruhe, Karlsruhe
 34. Patrick K, Norman GJ, Calfas KJ et al (2004) Diet, physical activity, and sedentary behaviors as risk factors for overweight in adolescence. *Arch Pediatr Adolesc Med* 158:385–390
 35. Prochaska JJ, Sallis JF, Long B (2001) A physical activity screening measure for use with adolescents in primary care. *Arch Pediatr Adolesc Med* 155:554–559
 36. Raczek J (2002) Entwicklungsveränderungen der motorischen Leistungsfähigkeit der Schuljugend in drei Jahrzehnten (1965–1995). *Sportwissenschaft* 32:201–216
 37. Romahn N (2008) Körperlich-sportliche Aktivität von Kindern und Jugendlichen in Deutschland: Eine repräsentative Befragung mit Kindern und Jugendlichen im Alter von 4–17 Jahren. Ph.D. dissertation. Universität Karlsruhe, Karlsruhe
 38. Rost R, Hollmann W (1982) Belastungsuntersuchungen in der Praxis. Grundlagen, Technik und Interpretation ergometrischer Untersuchungsverfahren. Thieme, Stuttgart
 39. Rowland TW (2001) The role of physical activity and fitness in children in the prevention of adult cardiovascular disease. *Prog Pediatr Cardiol* 12:199–203
 40. Sallis JF, Prochaska JJ, Taylor WC (2000) A review of correlates of physical activity of children and adolescents. *Med Sci Sports Exerc* 32:963–975
 41. Schilling F (1974) Körperkoordinationstest für Kinder. KTK. Manual. Beltz Test, Weinheim
 42. Schilling F, Baedke D (1980) Screening Test für den motorischen Bereich bei der Einschulung. *Motorik* 3:84
 43. Schmitz KH, Lytle LA, Phillips GA et al (2002) Psychosocial correlates of physical activity and sedentary leisure habits in young adolescents: the Teens Eating for Energy and Nutrition at School Study. *Prev Med* 34:266–278
 44. Schneider FJ (1986) Der neue gesundheitsorientierte Fitnesstest der USA und die Ergebnisse einer Untersuchung deutscher Kinder. *Sportunterricht* 35:173–181
 45. Simons J, Renson R (1982) Evaluation of motor fitness. Report of the European research seminar on the evaluation of motor fitness. Institute of Physical Education, Leuven
 46. Stemmler R (1953) Leistungen und Leistungsgrundwerte unserer Schüler. Volk und Wissen, Berlin
 47. Telama R (1982) Motor fitness test for Finnish schools. In: Simons J, Renson R (eds) *Evaluation of motor fitness. Report of the European research seminar on the evaluation of motor fitness*. Institute of Physical Education, Leuven, pp 169–198
 48. Thibault H, Contrand B, Saubusse E et al (2010) Risk factors for overweight and obesity in French adolescents: physical activity, sedentary behavior and parental characteristics. *Nutrition* 26:192–200
 49. Thompson AM, McHugh TL, Blanchard CM et al (2009) Physical activity of children and youth in Nova Scotia from 2001/02 and 2005/06. *Prev Med* 49:407–409
 50. Trost SG, Pate RR, Sallis JF et al (2002) Age and gender differences in objectively measured physical activity in youth. *Med Sci Sports Exerc* 34:350–355
 51. Trost SG, Rosenkranz RR, Dziewaltowski D (2008) Physical activity levels among children attending after-school programs. *Med Sci Sports Exerc* 40:622–629
 52. Ulmer J, Bös K (2000) Motorische Entwicklung salvadorianischer und deutscher Kinder. Ein Ländervergleich [Motor development of Children from El Salvador and Germany. A country comparison]. *Sportunterricht* 49:50–56
 53. van Mier H (2006) Developmental differences in drawing performance of the dominant and non-dominant hand in right-handed boys and girls. *Hum Mov Sci* 25:657–677
 54. Veitch J, Salmon J, Ball K (2010) Individual, social and physical environmental correlates of children's active free-play: a cross-sectional study. *Int J Behav Nutr Phys Act* 7:11
 55. Weineck J, Köstermeyer G, Sönnichsen A (1997) PEP eine Studie zur Präventionserziehung – Teil 1: Zum motorischen Leistungsvermögen von Schulanfängern. *Haltung und Bewegung* 17:5–16
 56. World Health Organisation Europe (WHO) (2004) Young people's health in context. *Health Behaviour in School-aged Children (HBSC) Study: international report from the 2001/2002 survey*. WHO, Copenhagen
 57. Wrotniak BH, Epstein LH, Dorn JM et al (2006) The relationship between motor proficiency and physical activity in children. *Pediatrics* 118:e1758–e1765

Copyright of European Journal of Pediatrics is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.



Review

Gender Differences in Fundamental Motor Skills Proficiency in Children Aged 3–6 Years: A Systematic Review and Meta-Analysis

Yunfei Zheng ¹, Weibing Ye ² , Mallikarjuna Korivi ² , Yubo Liu ^{2,*} and Feng Hong ^{3,*}¹ College of Physical Education and Health Sciences, Zhejiang Normal University, Jinhua 321004, China; zyf123@zjnu.edu.cn² Institute of Human Movement and Sports Engineering, Zhejiang Normal University, Jinhua 321004, China; ywbls@zjnu.cn (W.Y.); mallik.k5@gmail.com (M.K.)³ Department of Sports Operation and Management, Jinhua Polytechnic, Jinhua 321007, China

* Correspondence: liuyubo0124@outlook.com (Y.L.); fenghong0313@outlook.com (F.H.); Tel.: +86-183-2904-1023 (Y.L.)

Abstract: The age range of 3–6 years is considered as a critical period in developing and learning fundamental motor skills (FMS). To make the formulation of future FMS guidance programs more targeted, we examined gender differences in children's FMS proficiency using a meta-analysis. Structured electronic databases including PubMed, Scopus and Web of Science were systematically searched using key terms, and the Joanna Briggs Institute (JBI) was used to assess the quality of included literature. Finally, 38 articles (39 studies) met the pre-specified inclusion criteria. The results showed that boys had higher proficiency in total FMS and object control skills than girls (SMD = 0.17 (95% CI 0.03, 0.31), $p = 0.02$; SMD = 0.48 (95% CI 0.38, 0.58), $p < 0.00001$), and gender differences in locomotor skill proficiency approached significance, trending in favor of girls (SMD = -0.07 (95% CI $-0.15, 0.01$), $p = 0.09$, $I^2 = 66\%$). Meta-regression shows that age is associated with gender differences in object control skills ($p < 0.05$). In addition, through subgroup analysis, we found that boys' advantage in object control skills increased with age (3 years: SMD = 0.27 (95% CI 0.00, 0.54), $p < 0.00001$; 4 years: SMD = 0.58 (95% CI 0.38, 0.77), $p < 0.00001$; 5 years: SMD = 0.59 (95% CI 0.31, 0.88), $p < 0.00001$; 6 years: SMD = 0.81 (95% CI 0.61, 1.01), $p < 0.00001$). In this meta-analysis, we found gender differences in FMS levels in children aged 3–6 years. Notably, gender differences in skill proficiency in object control were influenced by age. We recommend focusing on and developing girls' object control skills starting at age 3.

Keywords: motor skills; child; Test of Gross Motor Development; sex differences

Citation: Zheng, Y.; Ye, W.; Korivi, M.; Liu, Y.; Hong, F. Gender Differences in Fundamental Motor Skills Proficiency in Children Aged 3–6 Years: A Systematic Review and Meta-Analysis. *Int. J. Environ. Res. Public Health* **2022**, *19*, 8318. <https://doi.org/10.3390/ijerph19148318>

Academic Editors: Clemens Drenowatz and Klaus Greier

Received: 28 May 2022

Accepted: 6 July 2022

Published: 7 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Regular participation in physical activity (PA) has potential benefits for children to improve obesity [1,2], bone health [3], psychological health [4] and cognitive function [5,6]. However, children's physical activity levels worldwide are not positive. A study comparing physical activity behaviors of children from 15 countries found that PA behavioral indicator scores were generally low [7]. Studies have found a positive correlation between children's fundamental motor skills (FMS) and PA, and FMS have been identified as a potential mechanism for the development of PA [8–11]. FMS refer to the basic abilities and skills for children to perform a series of organized basic movements, and they includes locomotor skills (e.g., running, jumping, sliding, etc.) and object control skills (e.g., hitting, catching ball, kicking, etc.) [12]. FMS play a vital role in using more professional and complex skills in playing, games and sports [13,14]. The learning and mastery of FMS play an important role in the healthy development of children. A previous study concluded that FMS in children were significantly associated with health-related fitness (HRF) components (body composition, muscular strength, muscular endurance, cardiovascular endurance) and that the effect increases with age [15]. Despite the many benefits of FMS, a recent systematic

review showed that there is still much room for improvement in FMS globally in children of all ages within the range of 3–10 years [16].

It has been well established that FMS are crucial to a child's development. When children are provided with few or no opportunities to achieve appropriate FMS levels, they are at risk of suffering from slowed motor development, thus limiting their chances for successful participation in an active and healthy sports culture [8]. Given the above, it seems crucial to improve children's FMS. A recently published study protocol presents detailed experimental designs to investigate the effects of different physical activity interventions on FMS in children [17]. However, to meet the physical developmental needs of children, exercise programs should also be tailored to their unique developmental needs, so it is important to understand gender characteristics in FMS.

However, no unified conclusion has been reached on whether there are gender differences in FMS proficiency in children. Several studies have found gender differences in FMS in children [16,18–20], with boys having higher proficiency in object control skills than girls [16,18,19,21,22]. In contrast, boys and girls have been found to have similar locomotor skills proficiency [18,19,21]. Pieces of evidence suggest gender differences in locomotor skills proficiency in children, with girls showing higher proficiency [20,22]. These inconsistent results may be clarified via a meta-analysis. To the best of our knowledge, no study has investigated age variation points for gender differences in motor skills. Hence, the main aim of the present study was to systematically review and provide a meta-analysis of the gender differences in FMS, locomotor skills and object control skills in children aged 3–6 years. Secondly, this study aims to investigate the age pattern of gender differences in motor skill proficiency by meta-regression analysis and determine the age inflection points at which gender differences emerge.

2. Materials and Methods

The systematic review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [23], and the study was registered with PROSPERO (CRD42021281160).

2.1. Eligibility Criteria

Articles assessing proficiency in FMS, locomotor skills and object control skills in children aged 3 to 6 years were included in this systematic review and meta-analysis. Studies were selected if they met the following criteria: (1) children aged 3–6 years; (2) scores reported by age and gender, if the study is an intervention study, with a baseline data report required; (3) results assessed using the Test of Gross Motor Development scale (TGMD), including modified versions; and (4) outcome measures reported using raw scores.

We excluded studies that met any of the following criteria: (1) review articles, conference abstracts or books; (2) participants not being assessed simultaneously according to age and gender; (3) inclusion of special populations such as those with disabilities or diseases; and (4) studies not published in English. We calculated the pooled effect size through meta-analysis.

Inclusion and exclusion criteria were formulated and applied by two reviewers (Y.Z. and M.K.) independently, and in the case of disagreement, they were confirmed by another reviewer (W.Y.).

2.2. Search Strategy

The literature search was conducted until 10 July 2021 by two independent reviewers (Y.Z. and Y.L.) with the following electronic databases: PubMed, Scopus and Web of Science (Supplementary Materials). Search terms included "child" OR "children" OR "preschoolers" OR "boy" OR "girl" AND "fundamental movement skills" OR "motor skills" OR "motor development" OR "gross motor" AND "TGMD" OR "Test of Gross Motor Development". Specific search strategies can be found in the Supplementary Materials. In addition, we manually searched reference lists in some key previously published studies and reviews.

2.3. Data Extraction and Quality Assessment

Data were independently extracted by two reviewers (W.Y. and F.H.) and cross-checked, and controversial issues were discussed based on the original text to determine the final outcome. The extracted information included study characteristics (author name, publication time, country, study type), participant characteristics (environment, age, gender, sample), measurement information (outcome measures, outcome indicators, outcome), etc.

The two reviewers (M.K. and F.H.) independently assessed the quality of each included article using the Joanna Briggs Institute (JBI) quality appraisal checklists for cross-sectional, case-control, cohort studies, quasi-experimental and randomized control trials [24] (Supplementary Materials). If there were disagreements, we discussed these together until consensus was reached. The critical evaluation checklist for various research methods has 8 to 13 items, and positive responses were rated as “yes”. We identified studies with an overall positive response between 50% and 75% as studies with moderate quality, and studies with a positive response over 75% were considered to be of high quality.

2.4. Synthesis Methods

In this study, differences in FMS proficiency among boys and girls were compared using Review Manager version 5.4.1 from the Cochrane Assistance Network. The data in this paper are continuous variables, and the comprehensive effect index is the Standard Mean Difference (SMD) and its 95% CI. SMD was interpreted as very small (<0.2), small (0.2–0.5), moderate (0.5–0.8) and large (>0.8) [25], where $p < 0.05$ indicates significant differences between the genders. The I-squared (I^2) statistic was used to test the heterogeneity between studies. When $I^2 \leq 50$, there was no heterogeneity between studies, and a fixed-effects model was used for meta-analysis; when $I^2 > 50$, there was heterogeneity between studies, and a random-effects model was used for meta-analysis [26]. The pooled results showed heterogeneity between studies, which we addressed by meta-regression analysis and subgroup analysis. In addition, we sequentially excluded literature for sensitivity analysis, evaluated the stability of the combined results of the meta-analysis and verified the existence of publication bias in the included studies using Egger’s test.

When more than two subgroups needed to be merged in the research, the first two subgroups were merged first, then the third subgroup was merged, and so on. The merge formula is as follows [27]:

$$SD = \sqrt{\frac{(N_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2 + \frac{N_1 N_2}{(N_1 + N_2)}(M_1^2 + M_2^2 - 2M_1 M_2)}{N_1 + N_2 - 1}}$$

where SD is standard deviation; Group 1 sample size is N_1 , mean is M_1 , standard deviation is SD_1 ; and Group 2 sample size is N_2 , mean is M_2 , and standard deviation is SD_2 .

3. Results

3.1. Search Results

A total of 2543 articles were retrieved through PubMed, Scopus and Web of Science databases, and 68 articles were obtained from reference lists of previously published studies and reviews. In the beginning, 623 duplicate records were removed from the 2543 articles. Of the remaining 1920 studies, 1688 irrelevant articles were excluded by reading the titles and abstracts, resulting in 232 items. After 15 articles that were not found were excluded, 217 articles were finally reviewed by full-text reading. From this stage, 44 articles were excluded due to age mismatch, 98 articles were removed because of insufficient information, 7 articles were excluded as they used other assessment tools, and 32 records were excluded due to other outcome measures. Ultimately, 36 articles met the inclusion criteria. In addition, 68 articles retrieved from the reference list were screened layer by layer, and 2 articles were finally included for the analysis. The detailed flow chart of the literature search, screening and selection is shown in Figure 1.

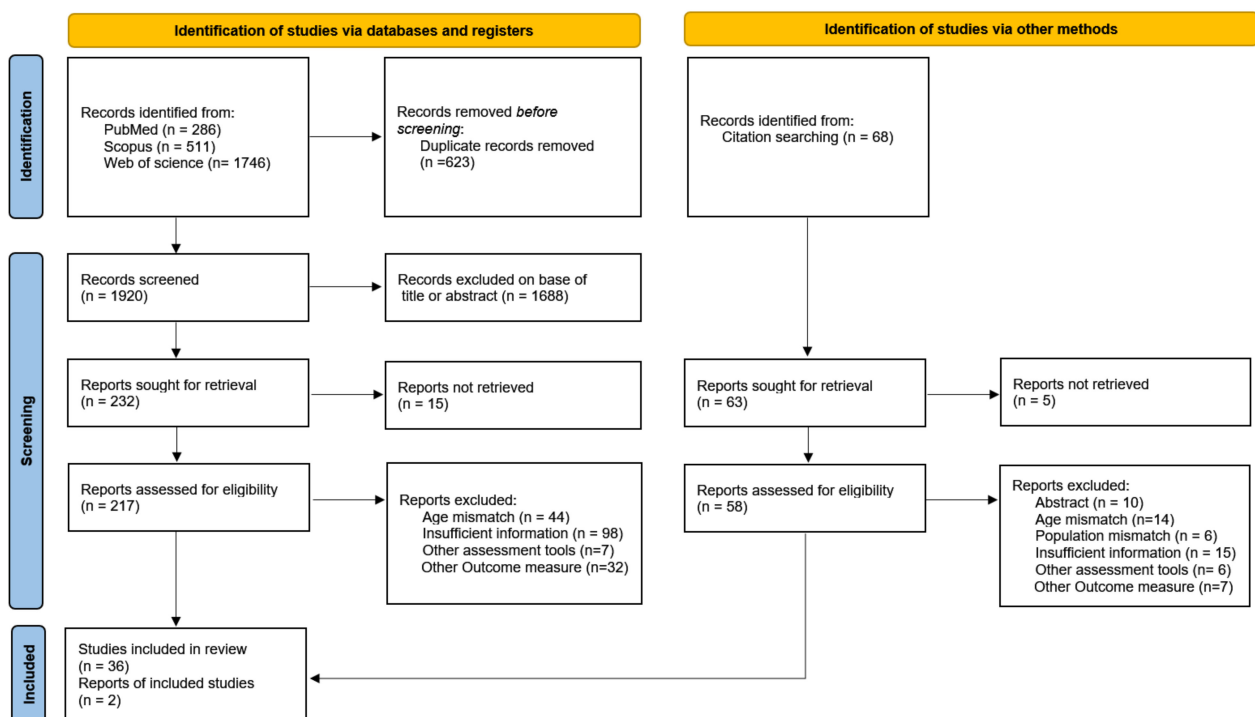


Figure 1. Flowchart of the study selection according to the Preferred Reporting Items for the Systematic Reviews and Meta-Analysis (PRISMA 2020) method.

3.2. Characteristics of Included Studies and Quality Assessment

After the inclusion and exclusion criteria had been applied, a total of 38 articles (39 studies) were included in the systematic review and meta-analysis. These studies were carried out in 19 different countries as follows: Australia [28], Belgium [29], Brazil [30–34], Britain [35–37], China [38–43], Croatia [44], Germany [45], Iran [46], Indonesia [47,48], Ireland [49,50], Japan [51], Korea [52], Myanmar [53], Poland [54,55], Portugal [56,57], Puerto Rico [58], Singapore [59], South Africa [60,61] and the USA [29,62–65]. A total of 2598 children participated in the FMS assessment, 8837 children participated in the locomotor skills assessment, and 8394 children participated in the object control skills assessment, ranging in age from 3 to 6 years old. A total of 33 studies were cross-sectional studies [28–38,40,43–45,47–63,65], one was a case-control study [39], one was a cohort study [42], two were quasi-experimental studies [46,64], and one was a randomized controlled study [41]. There were 2 studies that used TGMD-1 [44,58], 30 that used TGMD-2 [28–32,35–41,43,46,47,50–57,59–64], 1 that used the modified version of TGMD-2 [33], 5 that used TGMD-3 [42,45,48,49,65], and 1 that used the modified version of TGMD-3 [34]. Based on the JBI quality evaluation criteria, the quality scores of the included studies ranged from 60% to 100%. Most of the included articles are of high quality, a few are of medium quality, and none of them are of low quality. Test locations, test items, score reports and quality assessments are shown in Table 1.

Table 1. Characteristics of the included studies.

Author	Country	Design	Sample		Scale	Test Items	Outcome	Assessment
			Age	N (Boy/Girl)				
Alessandro et al. 2018 [30]	Brazil	CS	5–6	158 (82/76)	TGMD-2	LM, OC, FMS	FMS: ← LM: ↔ OC: ←	7
Aponte et al. 1990 [58]	Puerto Rico	CS	5–6	200 (102/98)	TGMD-1	FMS	FMS: ←	5
Aye et al. 2017 [53]	Myanmar	CS	5	472 (237/235)	TGMD-2	LM, OC	LM: ↔ OC: ←	7
Aye et al. 2018 [51]	Japan	CS	5	60 (34/26)	TGMD-2	LM, OC	LM: → OC: ←	7
Bakhtiar 2014 [47]	Indonesia	CS	6	67 (28/39)	TGMD-2	LM, OC	LM: ↔ OC: ↔	6
Behan et al. 2019 [49]	Ireland	CS	5–6	FMS: 357 (200/157) LM: 360 (202/158) OC: 359 (200/159)	TGMD-3	LM, OC, FMS	FMS: ↔ LM: → OC: ←	6
Bolger et al. 2017 [50]	Ireland	CS	6	102 (52/50)	TGMD-2	LM, OC	LM: → OC: ←	7
Brian et al. 2018 [29]	Belgium/USA	CS	4–5	Belgium: 170 (97/73) USA: 156 (66/90)	TGMD-2	LM, OC	Belgium: LM: ↔ OC: ← USA: LM: ↔ OC: ←	6
Brian et al. 2019 [62]	USA	CS	3–6	580 (284/296)	TGMD-2	LM, OC	LM: ↔ OC: ←	7
Capio et al. 2021 [38]	China	CS	4–6	230 (109/121)	TGMD-2	LM, OC	LM: ↔ OC: ←	6
Cheung et al. 2020 [39]	China	CC	4–6	295 (162/133)	TGMD-2	LM, OC, FMS	FMS: ↔ LM: ↔ OC: ↔	10
Cliff et al. 2009 [28]	Australia	CS	3–5	46 (25/21)	TGMD-2	LM, OC	LM: → OC: ↔	6
Famelia et al. 2018 [48]	Indonesia	CS	3–6	66 (30/36)	TGMD-3	LM, OC	LM: ↔ OC: ←	7
Freitas et al. 2018 [56]	Portugal	CS	3–6	314 (155/159)	TGMD-2	LM, OC	LM: ↔ OC: ←	8
Hall et al. 2018 [35]	Britain	CS	3–5	166 (91/75)	TGMD-2	LM, OC, FMS	FMS: ↔ LM: ↔ OC: ↔	7
Hall et al. 2019 [36]	Britain	CS	4–6	38 (24/14)	TGMD-2	LM, OC, FMS	FMS: ↔ LM: ↔ OC: ←	8
Henrique et al. 2020 [31]	Brazil	CS	3–5	472 (248/224)	TGMD-2	LM, OC	LM: ↔ OC: ←	7
Jiang et al. 2018 [40]	China	CS	3–6	60 (30/30)	TGMD-2	LM, OC, FMS	FMS: ↔ LM: ↔ OC: ↔	6
Kim et al. 2016 [52]	Korean	CS	5–6	216 (102/114)	TGMD-2	LM, OC	LM: ↔ OC: ←	5
Kit et al. 2017 [63]	United States	CS	3–5	LM: 330 (167/163) OC: 338 (170/168)	TGMD-2	LM, OC	LM: → OC: ←	7
Korbecki et al. 2017 [54]	Poland	CS	6	64 (35/29)	TGMD-2	LM, OC	LM: ↔ OC: ←	5
Kordi et al. 2012 [46]	Iran	QE	4–6	147 (75/72)	TGMD-2	LM, OC	LM: ↔ OC: ↔	7
Lopes et al. 2017 [57]	Portugal	CS	5–6	57 (26/31)	TGMD-2	LM, OC	LM: ↔ OC: ←	8
Mukherjee et al. 2017 [59]	Singapore	CS	6	95 (50/45)	TGMD-2	LM, OC	LM: ↔ OC: ←	6
Nikolić et al. 2016 [44]	Croatia	CS	4–4.5	67 (34/33)	TGMD-1	LM, OC, FMS	FMS: → LM: → OC: ↔	5
Palmer et al. 2020 [64]	USA	QE	3.5–5	54 (27/27)	TGMD-2	LM, OC, FMS	FMS: ↔ LM: ↔ OC: ↔	8
Roscoe et al. 2019 [37]	Britain	CS	3–4	185 (97/81)	TGMD-2	LM, OC, FMS	FMS: ↔ LM: → OC: ↔	6
Saczuk et al. 2021 [55]	Poland	CS	5	441 (255/186)	TGMD-2	LM, OC	LM: → OC: ↔	5
Shi et al. 2020 [41]	China	RCT	5–6	43 (22/21)	TGMD-2	LM, OC, FMS	FMS: ↔ LM: ↔ OC: ↔	9
Soares et al. 2020 [32]	Brazil	CS	3–5	251 (127/124)	TGMD-2	LM, OC, FMS	FMS: ← LM: ↔ OC: ←	7
Tietjens et al. 2018 [45]	Germany	CS	3–6	27 (11/16)	TGMD-3	LM, OC	LM: ↔ OC: ←	6
Tomaz et al. 2019 (1) [60]	South African	CS	3–6	259 (130/129)	TGMD-2	LM, OC, FMS	FMS: ← LM: ↔ OC: ←	7

Table 1. *Cont.*

Author	Country	Design	Sample		Scale	Test Items	Outcome	Assessment
			Age	N (Boy/Girl)				
Tomaz et al. 2019 (2) [61]	South African	CS	3–5	78 (39/39)	TGMD-2	LM, OC, FMS	FMS: ↔ LM: ↔ OC: ←	8
Valentini et al. 2012 [33]	Brazil	CS	3–6	LM: 786 (394/392) OC: 796 (394/402)	TGMD-2-BR	LM, OC	LM: ↔ OC: ←	6
Valentini et al. 2017 [34]	Brazil	CS	3–6	281 (135/146)	TGMD-3-BR	LM, OC	LM: ↔ OC: ←	6
Wang et al. 2020 [42]	China	Cs	3–6	268 (126/142)	TGMD-3	LM, OC, FMS	FMS: ↔ LM: ↔ OC: ↔	8
Webster et al. 2019 [65]	USA	CS	3–4	126 (58/68)	TGMD-3	LM, OC, FMS	FMS: ← LM: ↔ OC: ←	7
Wong & Cheung 2006 [43]	China	CS	3–6	797 (424/373)	TGMD-2	LM, OC	LM: ↔ OC: ←	7

CS: cross-sectional; CC: case-control; QE: quasi-experimental; Cs: cohort study RCT: randomized control trial; TGMD: test of gross motor development; FMS: fundamental movement skills; LM: locomotor skill; OC: object control skill; ←: favors boys; →: favors girls; ↔: no difference.

3.3. Gender Difference in Total FMS

Sixteen studies assessed total FMS [30,32,35–37,39–42,44,49,58,60,61,64,65], including 1351 boys and 1247 girls. Figure 2 displays the forest plots of standardized mean differences and 95% CI for the total FMS score (16 studies) based on the random effects meta-analysis results. Significant differences favor boys vs. girls (SMD = 0.17 (95% CI 0.03, 0.31), $p = 0.02$, $I^2 = 64$).

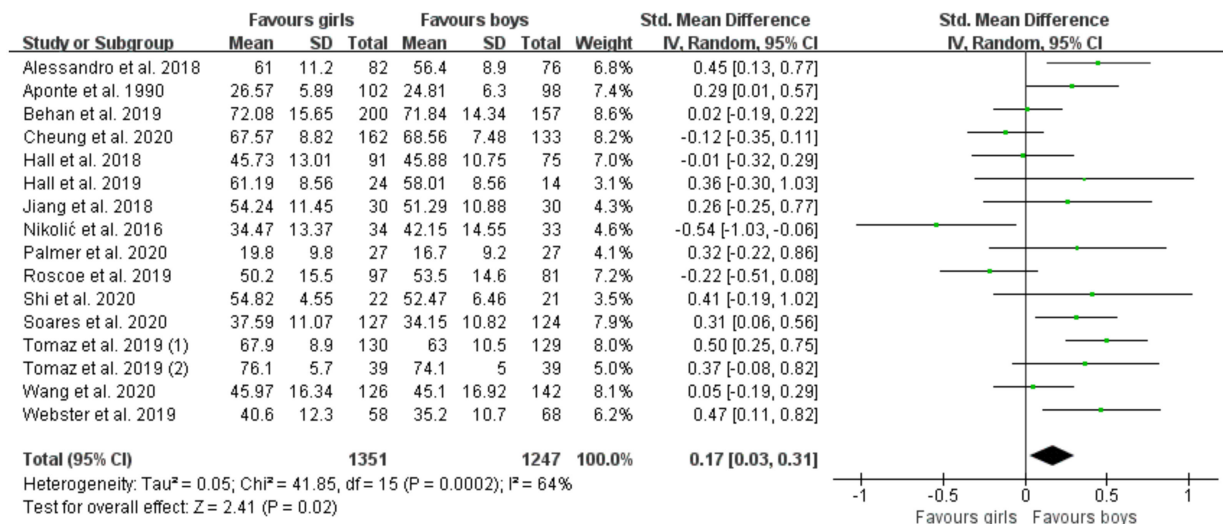


Figure 2. Forest plot of total FMS scores [30,32,35–37,39–42,44,49,58,60,61,64,65].

3.4. Gender Difference in Locomotor Skills

Thirty-seven articles (thirty-eight studies) assessed proficiency in locomotor skills [28–57,59–65], including 4290 boys and 4087 girls. Figure 3 displays the forest plots of standardized mean differences and 95% CI for the locomotor skills score (38 studies) based on the random effects meta-analysis results. Gender differences in locomotor skill proficiency approached significance, trending in favor of girls (SMD = -0.07 (95% CI $-0.15, 0.01$), $p = 0.09$, $I^2 = 66\%$).

3.5. Gender Difference in Object Control Skills

Thirty-seven articles (thirty-eight studies) assessed proficiency in object control skills [28–57,59–65], including 4291 boys and 4103 girls. Figure 4 displays forest plots of the standardized mean differences and 95% CI for the object control skills score (38 studies) based on the random effects meta-analysis results. Significant differences were found, favoring boys vs. girls (SMD = 0.48 (95% CI 0.38, 0.58), $p < 0.00001$). Meta-regression displays that age is associated with gender differences in object control skills ($p < 0.05$). To further explore the effect of age, we divided studies with age-specific assessments into a 3 year-old group, a 4 year-old group, a 5 year-old group and a 6 year-old group. In subgroup analyses (Figure 5), we found marginally significant results favoring boys vs. girls in children aged 3 (SMD = 0.27 (95% CI 0.00, 0.54), $p = 0.05$) and significant results favoring boys vs. girls aged 4, 5 and 6 years (SMD = 0.58 (95% CI 0.38, 0.77), $p < 0.00001$; SMD = 0.59 (95% CI 0.31, 0.88), $p < 0.00001$; SMD = 0.81 (95% CI 0.61, 1.01), $p < 0.00001$), which increased with age.

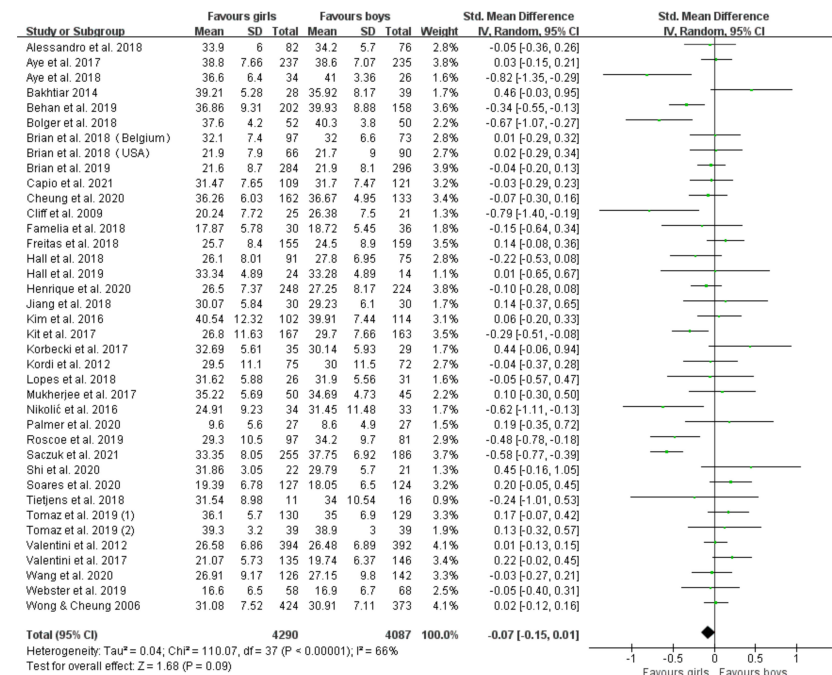
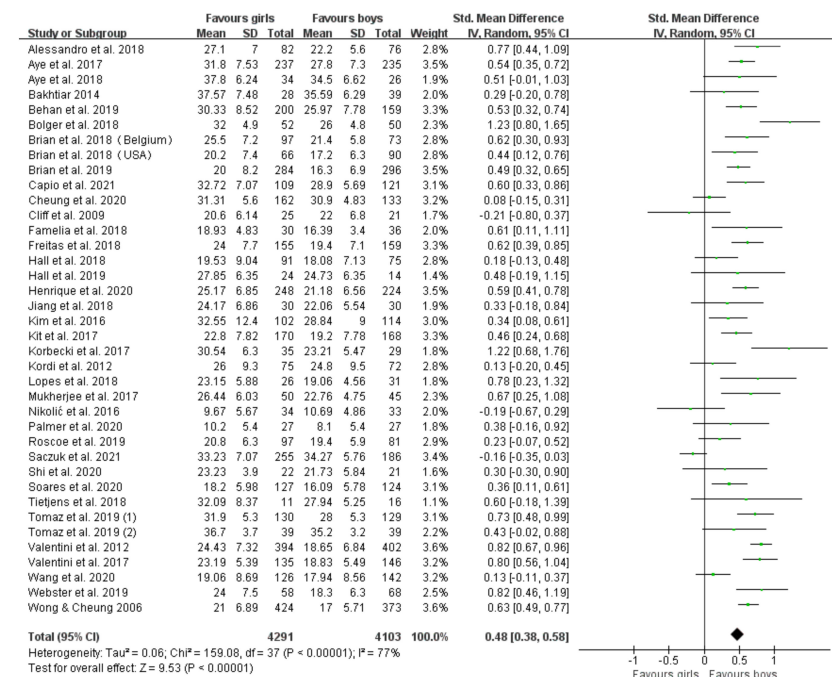
3.6. Sensitivity Analysis and Publication Bias

After excluding 39 studies one by one, it was found that there was no significant change in the magnitude or direction of differences in the proficiency of children of different genders in terms of FMS, locomotor skills or object control skills.

Egger's test was used to assess publication bias in FMS, locomotor skills and object control skills. The results showed that none of the studies included in the above review had publication bias ($p > 0.05$), as shown in Table 2.

Table 2. The result of publication bias estimation.

Item	Coef.	Std. Err.	t	p > t	95% Conf. Interval	
FMS	0.8661726	1.340221	0.65	0.529	−2.008316	3.740662
LM	−0.2382333	0.6962142	−0.34	0.734	−1.650221	1.173755
OC	−0.4231878	0.8376626	−0.51	0.616	−2.122046	1.275671

**Figure 3.** Forest plot of locomotor skills scores [28–57,59–65].**Figure 4.** Forest plot of object control skills scores [28–57,59–65].

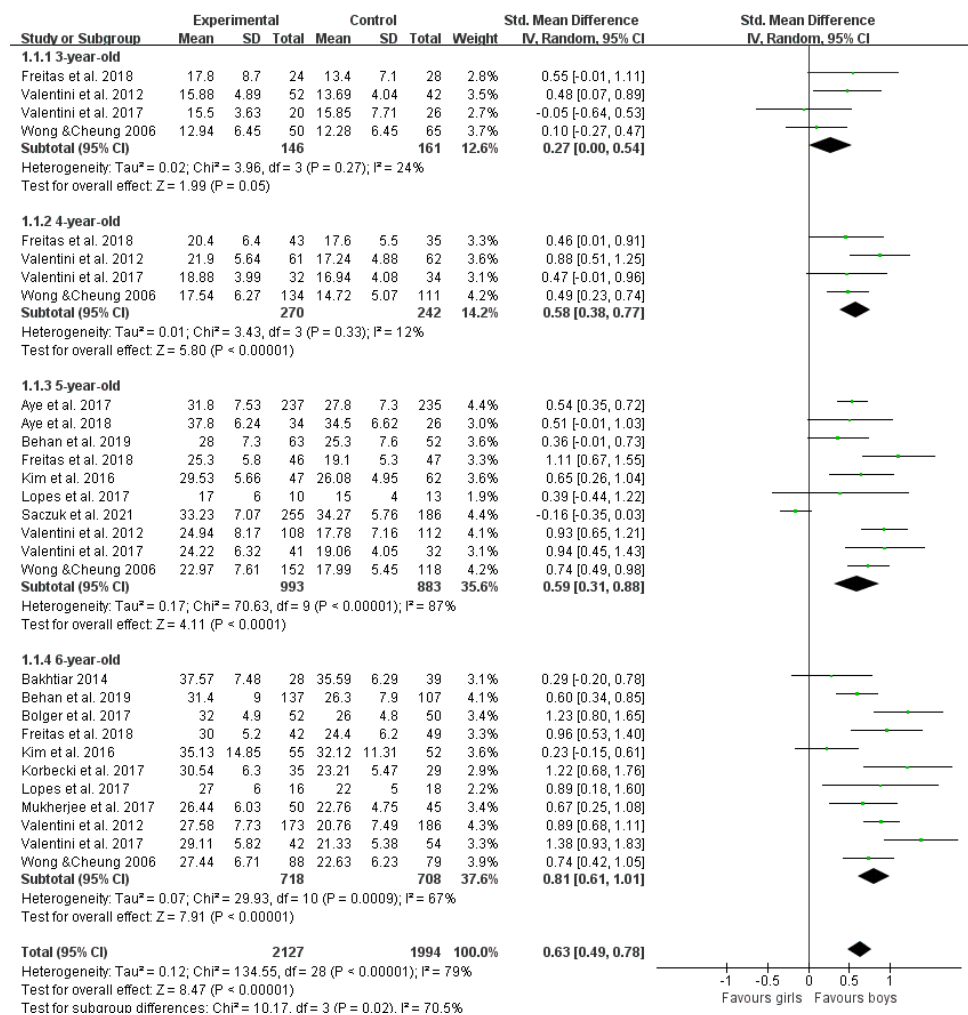


Figure 5. Forest plot of object control skills (age subgroups) [28–57,59–65].

4. Discussion

This systematic review and meta-analysis aggregated studies from Asia (China [38–43], Iran [46], Indonesia [47,48], Korean [52], Myanmar [53], Japan [51] and Singapore [59]), Africa (South Africa [60,61]), Europe (Belgium [29], Britain [35–37], Croatia [44], Germany [45], Ireland [49,50], Poland [54,55] and Portugal [56,57]), North America (the United States (US) [29,62–65] and Puerto Rico), Oceania (Australia [28]) and South America (Brazil [30–34]) and demonstrated gender differences in FMS proficiency in children aged 3–6 years. Combined results show that boys are more proficient than girls in total FMS proficiency. From the two dimensions of proficiency in locomotor skills and in object control skills, marginally significant differences were found favoring girls, and significant differences were found favoring boys in object control skills.

Differences in proficiency in object control skills between boys and girls seem to take some cues from biology. A study reported that boys are more likely to use finely segmented pelvic–torso–shoulder rotation when throwing [66]. Young explained the differences in human throwing and hitting behavior from an evolutionary perspective. Early humans made a living by throwing stones and swinging clubs. Women invested more resources into reproduction, and men were more likely to be hunters and warriors. These kinds of patterns are inherited through natural selection [67]. A previous study speculates that mature throwing is more likely an innate skill whose development is biologically determined and somewhat difficult to be influenced by nurture, and the same may be true of striking [68]. Sociological factors and behavior habits may also contribute to gender differences in proficiency in object control. Physical education programs are important

for the development of FMS in preschoolers. Research shows that structured physical activity lessons can improve children's FMS [69]. A meta-analysis shows that three or more teacher-led physical activity sessions per week significantly improved FMS [11]. Furthermore, studies have shown a correlation between FMS proficiency and physical activity levels in children. A study using TGMD-2 and accelerometers measured data on FMS and physical activity in kindergarteners and found a positive relationship between object control skills and moderate-intensity physical activity (MVPA) [70]. However, it has been shown that girls are significantly less likely to participate in physical activity than boys during the preschool years, especially at moderate to high intensity [71]. A systematic review including 10,316 children aged 3–6 years (5236 boys and 5080 girls) demonstrated that boys were more physically active than girls [72]. A survey in Norway showed that among children aged 3–4, only 32% of girls and 67% of boys were able to achieve the recommended 60 min of moderate to vigorous physical activity per day [73]. Therefore, different levels of physical activity may be responsible for the gender differences in object control scores. In addition, differences in exercise content may also contribute to gender differences in object control. A cross-sectional study from Japan showed that 5 year-old boys had significantly higher raw scores in terms of object control than girls of the same age (37.8 ± 6.24 vs. 34.5 ± 6.62 , respectively), which is consistent with our findings, and the difference is mainly reflected in hitting, kicking and throwing [51]. A study in Australia showed that girls opted for dance and aerobic exercises far more often than boys [74]. A study found that Taiwanese girls prefer to play hopscotch, balance beam and house, while boys prefer ball games and slapstick games [75]. Previous studies have indicated that girls tend to lack opportunities to practice ball games, while boys generally spend more time participating in these games [76,77], which may also be related to parental educational attitudes [78]. In addition, an interesting study in Canada showed that 5 year-old girls' perception of physical ability was related to their proficiency in locomotor skills, but not to object control skills, which may be because girls do not value object control skills [79].

Object control skills are more important than locomotor skills in childhood and continue to affect adolescence [44]. Evidence shows that gender disparities are reduced if girls have the same opportunities for mentoring, feedback, practice and encouragement [64]. Our meta-regression analysis revealed that age was the main factor influencing differences in proficiency in object control skills between boys and girls. Using subgroup analysis to further explore the effect of age, we found that gender differences in children's proficiency in object control skill tend to be significant at age 3, and the advantage tends to favor boys. The difference is significant at the age of 4, and the advantage of boys begins to gradually increase with age, reaching a maximum at the age of 6. We recommend that parents and teachers should start paying attention to children's movements when they are 3 years old and consciously guide children's sports participation types; in particular, girls are encouraged to participate in ball games. Scholars should comprehensively consider the growth and development patterns, types of exercise and professional guidance of boys and girls when studying FMS guidance plans for children.

This meta-analysis provides evidence for gender differences in FMS proficiency in children aged 3–6 years, but some limitations should be considered. First, there are fewer articles and a smaller sample size for children aged 3 and 4 years, which requires more data to confirm. Second, our study only included children aged 3–6 years, and gender differences in FMS in children of other ages are also an important topic. Third, due to the limited number of articles, the study could not be specific to each item in the TGMD subscale (e.g., running, jumping, dribbling, etc.). Studies on specific TGMD items will therefore also be an interesting and useful topic as the number of high-quality studies increases. Finally, because TGMD is the most common tool for measuring FMS proficiency in educational, clinical and research settings, our study only included articles using TGMD or any modified version (TGMD-2 or TGMD-3), but this may have led to inconsistent results. Currently, assessments of children's FMS competencies are primarily conducted through process-oriented and product-oriented approaches. TGMD is a process-oriented

assessment that examines children's motor performance on locomotor and object control tasks. We suggest that future research use product-oriented tools to further explore the gendered characteristics of FMS proficiency in young children.

5. Conclusions

Our findings demonstrated that there were gender differences in total FMS proficiency in children aged 3–6, with boys being more proficient than girls, and locomotor proficiency differences between gender approached significance, with a trend favoring girls. In the performance of proficiency in object control skills, boys were better than girls, and this difference gradually increased with age. We recommend focusing on and developing girls' object control skills starting at age 3.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijerph19148318/s1>.

Author Contributions: Conceptualization: Y.Z. and W.Y.; methodology: Y.Z. and Y.L.; software: Y.Z. and Y.L.; validation: M.K. and F.H.; formal analysis: Y.Z., W.Y. and Y.L.; investigation: Y.Z. and M.K.; resources: W.Y.; data curation: W.Y., Y.L. and F.H.; writing—original draft preparation: Y.Z.; writing—review and editing: M.K., Y.L. and F.H.; visualization: Y.Z.; supervision: W.Y.; project administration: Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the grant (KYH06Y21383) provided by Open Research Fund of College of Teacher Education, for the early development of infants and children aged 0–3 years old in China: A study on the characteristics and evaluation of physical and motor development.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are thankful to the Institute of Human Movement and Sports Engineering, Zhejiang Normal University for support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Psaltopoulou, T.; Tzanninis, S.; Ntanas-Stathopoulos, I.; Panotopoulos, G.; Kostopoulou, M.; Tzanninis, I.G.; Tsagianni, A.; Sergeantanis, T.N. Prevention and treatment of childhood and adolescent obesity: A systematic review of meta-analyses. *World J. Pediatr.* **2019**, *15*, 350–381. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Arhab, A.; Messerli-Bürge, N.; Kakebeeke, T.H.; Stülz, K.; Zysset, A.; Leeger-Aschmann, C.S.; Schmutz, E.A.; Meyer, A.H.; Munsch, S.; Kriemler, S.; et al. Association of physical activity with adiposity in preschoolers using different clinical adiposity measures: A cross-sectional study. *BMC Pediatr.* **2019**, *19*, 397. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Loprinzi, P.D.; Cardinal, B.J.; Loprinzi, K.L.; Lee, H. Benefits and environmental determinants of physical activity in children and adolescents. *Obes. Facts* **2012**, *5*, 597–610. [\[CrossRef\]](#)
4. Eime, R.M.; Young, J.A.; Harvey, J.T.; Charity, M.J.; Payne, W.R. A systematic review of the psychological and social benefits of participation in sport for children and adolescents: Informing development of a conceptual model of health through sport. *Int. J. Behav. Nutr. Phys. Act.* **2013**, *10*, 98. [\[CrossRef\]](#)
5. Martin, A.; Booth, J.N.; Laird, Y.; Sproule, J.; Reilly, J.J.; Saunders, D.H. Physical activity, diet and other behavioural interventions for improving cognition and school achievement in children and adolescents with obesity or overweight. *Cochrane Database Syst. Rev.* **2018**. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Donnelly, J.E.; Hillman, C.H.; Castelli, D.; Etnier, J.L.; Lee, S.; Tomporowski, P.; Lambourne, K.; Szabo-Reed, A.N. Physical activity, fitness, cognitive function, and academic achievement in children: A systematic review. *Med. Sci. Sports Exerc.* **2016**, *48*, 1197. [\[CrossRef\]](#)
7. Tremblay, M.S.; Gray, C.E.; Akinroye, K.; Harrington, D.M.; Katzmarzyk, P.T.; Lambert, E.V.; Prista, A. Physical activity of children: A global matrix of grades comparing 15 countries. *J. Phys. Act. Health* **2014**, *11*, S113–S125. [\[CrossRef\]](#)
8. Stodden, D.F.; Goodway, J.D.; Langendorfer, S.J.; Robertson, M.A.; Rudisill, M.E.; Garcia, C.; Garcia, L.E. A developmental perspective on the role of motor skill competence in physical activity: An emergent relationship. *Quest* **2008**, *60*, 290–306. [\[CrossRef\]](#)
9. Lubans, D.R.; Morgan, P.J.; Cliff, D.P.; Barnett, L.M.; Okely, A.D. Fundamental movement skills in children and adolescents. *Sports Med.* **2010**, *40*, 1019–1035. [\[CrossRef\]](#)

10. Williams, H.G.; Pfeiffer, K.A.; O'Neill, J.R.; Dowda, M.; McIver, K.L.; Brown, W.H.; Pate, R.R. Motor skill performance and physical activity in preschool children. *Obesity* **2008**, *16*, 1421–1426. [\[CrossRef\]](#)
11. Engel, A.C.; Broderick, C.R.; van Doorn, N.; Hardy, L.L.; Parmenter, B.J. Exploring the relationship between fundamental motor skill interventions and physical activity levels in children: A systematic review and meta-analysis. *Sports Med.* **2018**, *48*, 1845–1857. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Wick, K.; Leeger-Aschmann, C.S.; Monn, N.D.; Radtke, T.; Ott, L.V.; Rebholz, C.E.; Cruz, S.; Gerber, N.; Schmutz, E.A.; Puder, J.J. Interventions to promote fundamental movement skills in childcare and kindergarten: A systematic review and meta-analysis. *Sports Med.* **2017**, *47*, 2045–2068. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Logan, S.W.; Ross, S.M.; Chee, K.; Stodden, D.F.; Robinson, L.E. Fundamental motor skills: A systematic review of terminology. *J. Sports Sci.* **2018**, *36*, 781–796. [\[CrossRef\]](#)
14. Zhang, L.; Cheung, P. Making a difference in PE lessons: Using a low organized games approach to teach fundamental motor skills in China. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4618. [\[CrossRef\]](#)
15. Behan, S.; Belton, S.; Peers, C.; O'Connor, N.E.; Issartel, J. Exploring the relationships between fundamental movement skills and health related fitness components in children. *Eur. J. Sport Sci.* **2022**, *22*, 171–181. [\[CrossRef\]](#)
16. Bolger, L.E.; Bolger, L.A.; O'Neill, C.; Coughlan, E.; O'Brien, W.; Lacey, S.; Burns, C.; Bardid, F. Global levels of fundamental motor skills in children: A systematic review. *J. Sports Sci.* **2021**, *39*, 717–753. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Wang, G.; Zi, Y.; Li, B.; Su, S.; Sun, L.; Wang, F.; Ren, C.; Liu, Y. The Effect of Physical Exercise on Fundamental Movement Skills and Physical Fitness among Preschool Children: Study Protocol for a Cluster-Randomized Controlled Trial. *Int. J. Environ. Res. Public Health* **2022**, *19*, 6331. [\[CrossRef\]](#)
18. Robinson, L.E. The relationship between perceived physical competence and fundamental motor skills in preschool children. *Child Care Health Dev.* **2011**, *37*, 589–596. [\[CrossRef\]](#)
19. Robinson, L.E.; Wadsworth, D.D.; Peoples, C.M. Correlates of school-day physical activity in preschool students. *Res. Q. Exerc. Sport* **2012**, *83*, 20–26. [\[CrossRef\]](#)
20. Niemistö, D.; Finni, T.; Cantell, M.; Korhonen, E.; Sääkslahti, A. Individual, family, and environmental correlates of motor competence in young children: Regression model analysis of data obtained from two motor tests. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2548. [\[CrossRef\]](#)
21. Barnett, L.M.; Van Beurden, E.; Morgan, P.J.; Brooks, L.O.; Beard, J.R. Gender differences in motor skill proficiency from childhood to adolescence: A longitudinal study. *Res. Q. Exerc. Sport* **2010**, *81*, 162–170. [\[CrossRef\]](#)
22. Temple, V.A.; Crane, J.R.; Brown, A.; Williams, B.-L.; Bell, R.I. Recreational activities and motor skills of children in kindergarten. *Phys. Educ. Sport Pedagog.* **2016**, *21*, 268–280. [\[CrossRef\]](#)
23. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Syst. Rev.* **2021**, *10*, 89. [\[CrossRef\]](#)
24. Moola, S.; Munn, Z.; Tufanaru, C.; Aromataris, E.; Mu, P.F. Chapter 7: Systematic Reviews of Etiology and Risk. In *JBIM Manual for Evidence Synthesis*; Aromataris, E., Munn, Z., Eds.; JBI: Adelaide, Australia, 2020.
25. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Academic Press: Cambridge, MA, USA, 1988.
26. Higgins, J.P.; Thompson, S.G.; Deeks, J.J.; Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **2003**, *327*, 557–560. [\[CrossRef\]](#)
27. Higgins, J.P.; Deeks, J.; Higgins, J.; Green, S. Chapter 7: Selecting studies and collecting data. In *Cochrane Handbook of Systematic Reviews of Interventions*; Version 5.1.0; Updated March 2011; Cochrane Collaboration: London, UK, 2011.
28. Cliff, D.P.; Okely, A.D.; Smith, L.M.; McKeen, K. Relationships between fundamental movement skills and objectively measured physical activity in preschool children. *Pediatric Exerc. Sci.* **2009**, *21*, 436–449. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Brian, A.; Bardid, F.; Barnett, L.M.; Deconinck, F.J.; Lenoir, M.; Goodway, J.D. Actual and perceived motor competence levels of Belgian and United States preschool children. *J. Mot. Learn. Dev.* **2018**, *6*, S320–S336. [\[CrossRef\]](#)
30. Ré, A.H.; Logan, S.W.; Cattuzzo, M.T.; Henrique, R.S.; Tudela, M.C.; Stodden, D.F. Comparison of motor competence levels on two assessments across childhood. *J. Sports Sci.* **2018**, *36*, 1–6. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Henrique, R.S.; Stodden, D.F.; Franssen, J.; Feitoza, A.H.; Ré, A.H.; Martins, C.M.; Dos Prazeres, T.M.; Cattuzzo, M.T. Is motor competence associated with the risk of central obesity in preschoolers? *Am. J. Hum. Biol.* **2020**, *32*, e23364. [\[CrossRef\]](#)
32. Soares, Í.A.A.; Martins, C.M.d.L.; Nobre, G.C.; Cattuzzo, M.T. Evidences of construct validity, criteria and validation of the motor competence assessment battery of tests in preschoolers. *J. Phys. Educ.* **2020**, *31*. [\[CrossRef\]](#)
33. Valentini, N.C. Validity and reliability of the TGMD-2 for Brazilian children. *J. Mot. Behav.* **2012**, *44*, 275–280. [\[CrossRef\]](#)
34. Valentini, N.C.; Zanella, L.W.; Webster, E.K. Test of Gross Motor Development—Third edition: Establishing content and construct validity for Brazilian children. *J. Mot. Learn. Dev.* **2017**, *5*, 15–28. [\[CrossRef\]](#)
35. Hall, C.J.; Eyre, E.L.; Oxford, S.W.; Duncan, M.J. Relationships between motor competence, physical activity, and obesity in British preschool aged children. *J. Funct. Morphol. Kinesiol.* **2018**, *3*, 57. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Hall, C.J.; Eyre, E.L.; Oxford, S.W.; Duncan, M.J. Does perception of motor competence mediate associations between motor competence and physical activity in early years children? *Sports* **2019**, *7*, 77. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Roscoe, C.M.; James, R.S.; Duncan, M.J. Accelerometer-based physical activity levels, fundamental movement skills and weight status in British preschool children from a deprived area. *Eur. J. Pediatrics* **2019**, *178*, 1043–1052. [\[CrossRef\]](#)

38. Capio, C.M.; Eguia, K.F. Movement skills, perception, and physical activity of young children: A mediation analysis. *Pediatrics Int.* **2021**, *63*, 442–447. [\[CrossRef\]](#)
39. Cheung, P.; Zhang, L. Environment for Preschool Children to Learn Fundamental Motor Skills: The Role of Teaching Venue and Class Size. *Sustainability* **2020**, *12*, 9774. [\[CrossRef\]](#)
40. Jiang, G.-P.; Jiao, X.-B.; Wu, S.-K.; Ji, Z.-Q.; Liu, W.-T.; Chen, X.; Wang, H.-H. Balance, proprioception, and gross motor development of chinese children aged 3 to 6 years. *J. Mot. Behav.* **2018**, *50*, 343–352. [\[CrossRef\]](#)
41. Shi, K.; Sun, X.; Wang, Y.; Zha, P. Effects of gymnastics intervention on gross motor development in children aged 5 to 6 years: A randomized, controlled trial. *Med. Dello Sport* **2020**, *73*, 327–336. [\[CrossRef\]](#)
42. Wang, H.; Chen, Y.; Liu, J.; Sun, H.; Gao, W. A follow-up study of motor skill development and its determinants in preschool children from middle-income family. *BioMed Res. Int.* **2020**, *2020*, 6639341. [\[CrossRef\]](#)
43. Wong, A.K.Y.; Cheung, S.Y. Gross Motor Skills Performance of Hong Kong Chinese Children. *Asian J. Phys. Educ. Recreat.* **2006**, *12*, 23–29. [\[CrossRef\]](#)
44. Nikolić, I.; Mraković, S.; Kunješić, M. Gender differences of preschool children in fundamental movement skills. *Croat. J. Educ.* **2016**, *18*, 123–131. [\[CrossRef\]](#)
45. Tietjens, M.; Dreiskaemper, D.; Utesch, T.; Schott, N.; Barnett, L.M.; Hinkley, T. Pictorial scale of physical self-concept for younger children (P-PSC-C): A feasibility study. *J. Mot. Learn. Dev.* **2018**, *6*, S391–S402. [\[CrossRef\]](#)
46. Kordi, R.; Nourian, R.; Ghayour, M.; Kordi, M.; Younesian, A. Development and evaluation of a basic physical and sports activity program for preschool children in nursery schools in Iran: An interventional study. *Iran. J. Pediatrics* **2012**, *22*, 357. [\[CrossRef\]](#)
47. Bakhtiar, S. Fundamental motor skill among 6-year-old children in Padang, West Sumatera, Indonesia. *Asian Soc. Sci.* **2014**, *10*, 155–158. [\[CrossRef\]](#)
48. Famelia, R.; Tsuda, E.; Bakhtiar, S.; Goodway, J.D. Relationships among perceived and actual motor skill competence and physical activity in Indonesian preschoolers. *J. Mot. Learn. Dev.* **2018**, *6*, S403–S423. [\[CrossRef\]](#)
49. Behan, S.; Belton, S.; Peers, C.; O'Connor, N.E.; Issartel, J. Moving Well-Being Well: Investigating the maturation of fundamental movement skill proficiency across sex in Irish children aged five to twelve. *J. Sports Sci.* **2019**, *37*, 2604–2612. [\[CrossRef\]](#)
50. Bolger, L.E.; Bolger, L.A.; O'Neill, C.; Coughlan, E.; O'Brien, W.; Lacey, S.; Burns, C. Age and sex differences in fundamental movement skills among a cohort of Irish school children. *J. Mot. Learn. Dev.* **2018**, *6*, 81–100. [\[CrossRef\]](#)
51. Aye, T.; Kuramoto-Ahuja, T.; Sato, T.; Sadakiyo, K.; Watanabe, M.; Maruyama, H. Gross motor skill development of kindergarten children in Japan. *J. Phys. Ther. Sci.* **2018**, *30*, 711–715. [\[CrossRef\]](#)
52. Kim, C.-I.; Lee, K.-Y. The relationship between fundamental movement skills and body mass index in Korean preschool children. *Eur. Early Child. Educ. Res. J.* **2016**, *24*, 928–935. [\[CrossRef\]](#)
53. Aye, T.; Oo, K.S.; Khin, M.T.; Kuramoto-Ahuja, T.; Maruyama, H. Gross motor skill development of 5-year-old Kindergarten children in Myanmar. *J. Phys. Ther. Sci.* **2017**, *29*, 1772–1778. [\[CrossRef\]](#)
54. Korbecki, M.; Wawrzyniak, S.; Rokita, A. Fundamental movement skills of six-to seven-year-old children in the first grade of elementary school: A pilot study. *Balt. J. Health Phys. Act.* **2017**, *9*, 2. [\[CrossRef\]](#)
55. Saczuk, J.; Wasiluk, A. Assessment of the relationship between fitness abilities and motor skills of 5-year-olds by taking into account dimorphic differences. *J. Phys. Educ. Sport* **2021**, *21*, 115–121. [\[CrossRef\]](#)
56. Freitas, D.; Lausen, B.; Maia, J.; Gouveia, É.; Antunes, A.; Thomis, M.; Lefevre, J.; Malina, R. Skeletal maturation, fundamental motor skills, and motor performance in preschool children. *Scand. J. Med. Sci. Sports* **2018**, *28*, 2358–2368. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Lopes, V.P.; Saraiva, L.; Rodrigues, L.P. Reliability and construct validity of the test of gross motor development-2 in Portuguese children. *Int. J. Sport Exerc. Psychol.* **2018**, *16*, 250–260. [\[CrossRef\]](#)
58. Aponte, R.; French, R.; Sherrill, C. Motor development of Puerto Rican children: Cross-cultural perspectives. *Percept. Mot. Ski.* **1990**, *71*, 1200–1202. [\[CrossRef\]](#)
59. Mukherjee, S.; Ting Jamie, L.C.; Fong, L.H. Fundamental motor skill proficiency of 6-to 9-year-old Singaporean children. *Percept. Mot. Ski.* **2017**, *124*, 584–600. [\[CrossRef\]](#)
60. Tomaz, S.; Jones, R.A.; Hinkley, T.; Bernstein, S.; Twine, R.; Kahn, K.; Norris, S.A.; Draper, C.E. Gross motor skills of South African preschool-aged children across different income settings. *J. Sci. Med. Sport* **2019**, *22*, 689–694. [\[CrossRef\]](#)
61. Tomaz, S.A.; Pioreschi, A.; Watson, E.D.; McVeigh, J.A.; Rae, D.E.; Jones, R.A.; Draper, C.E. Body mass index, physical activity, sedentary behavior, sleep, and gross motor skill proficiency in preschool children from a low-to middle-income urban setting. *J. Phys. Act. Health* **2019**, *16*, 525–532. [\[CrossRef\]](#)
62. Brian, A.; Pennell, A.; Taunton, S.; Starrett, A.; Howard-Shaughnessy, C.; Goodway, J.D.; Wadsworth, D.; Rudisill, M.; Stodden, D. Motor competence levels and developmental delay in early childhood: A multicenter cross-sectional study conducted in the USA. *Sports Med.* **2019**, *49*, 1609–1618. [\[CrossRef\]](#)
63. Kit, B.K.; Akinbami, L.J.; Isfahani, N.S.; Ulrich, D.A. Gross motor development in children aged 3–5 years, United States 2012. *Matern. Child Health J.* **2017**, *21*, 1573–1580. [\[CrossRef\]](#)
64. Palmer, K.K.; Harkavy, D.; Rock, S.M.; Robinson, L.E. Boys and girls have similar gains in fundamental motor skills across a preschool motor skill intervention. *J. Mot. Learn. Dev.* **2020**, *8*, 569–579. [\[CrossRef\]](#)
65. Webster, E.K.; Martin, C.K.; Staiano, A.E. Fundamental motor skills, screen-time, and physical activity in preschoolers. *J. Sport Health Sci.* **2019**, *8*, 114–121. [\[CrossRef\]](#) [\[PubMed\]](#)

66. Butterfield, S.A.; Angell, R.M.; Mason, C.A. Age and sex differences in object control skills by children ages 5 to 14. *Percept. Mot. Ski.* **2012**, *114*, 261–274. [[CrossRef](#)]
67. Young, R.W. The ontogeny of throwing and striking. *Hum. Ontog. Int. J. Interdiscip. Dev. Res.* **2009**, *3*, 19–31. [[CrossRef](#)]
68. Angell, R.M.; Butterfield, S.A.; Tu, S.; Loovis, E.M.; Mason, C.A.; Nightingale, C.J. Children's throwing and striking: A longitudinal study. *J. Mot. Learn. Dev.* **2018**, *6*, 315–332. [[CrossRef](#)]
69. Jones, R.A.; Okely, A.D.; Hinkley, T.; Batterham, M.; Burke, C. Promoting gross motor skills and physical activity in childcare: A translational randomized controlled trial. *J. Sci. Med. Sport* **2016**, *19*, 744–749. [[CrossRef](#)] [[PubMed](#)]
70. Crane, J.R.; Naylor, P.J.; Cook, R.; Temple, V.A. Do Perceptions of Competence Mediate The Relationship Between Fundamental Motor Skill Proficiency and Physical Activity Levels of Children in Kindergarten? *J. Phys. Act. Health* **2015**, *12*, 954–961. [[CrossRef](#)]
71. Pate, R.R.; Pfeiffer, K.A.; Trost, S.G.; Ziegler, P.; Dowda, M. Physical activity among children attending preschools. *Pediatrics* **2004**, *114*, 1258–1263. [[CrossRef](#)]
72. Tucker, P. The physical activity levels of preschool-aged children: A systematic review. *Early Child. Res. Q.* **2008**, *23*, 547–558. [[CrossRef](#)]
73. Andersen, E.; Borch-Jenssen, J.; Øvreås, S.; Ellingsen, H.; Jørgensen, K.A.; Moser, T. Objectively measured physical activity level and sedentary behavior in Norwegian children during a week in preschool. *Prev. Med. Rep.* **2017**, *7*, 130–135. [[CrossRef](#)]
74. Dudley, D.A.; Cotton, W.G.; Peralta, L.R.; Winslade, M. Playground activities and gender variation in objectively measured physical activity intensity in Australian primary school children: A repeated measures study. *BMC Public Health* **2018**, *18*, 1101. [[CrossRef](#)] [[PubMed](#)]
75. Tsai, C.; Yang, S. Study on the appearance of childhood games. *J. Educ. Stud.* **2012**, *46*, 1–19.
76. Ogden, C.L.; Flegal, K.M.; Carroll, M.D.; Johnson, C.L. Prevalence and trends in overweight among US children and adolescents, 1999–2000. *JAMA* **2002**, *288*, 1728–1732. [[CrossRef](#)] [[PubMed](#)]
77. Thornton, M. *Life Span Motor Development*, 4th ed.; Physiotherapy Canada: Ottawa, ON, Canada, 2006; Volume 58, p. 240. [[CrossRef](#)]
78. Fagot, B.I.; Leinbach, M.D. The young child's gender schema: Environmental input, internal organization. *Child Dev.* **1989**, *60*, 663–672. [[CrossRef](#)] [[PubMed](#)]
79. LeGear, M.; Greyling, L.; Sloan, E.; Bell, R.I.; Williams, B.L.; Naylor, P.J.; Temple, V.A. A window of opportunity? Motor skills and perceptions of competence of children in kindergarten. *Int. J. Behav. Nutr. Phys. Act.* **2012**, *9*, 29. [[CrossRef](#)]

DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS

FIFTH EDITION
TEXT REVISION

DSM-5-TR™



Copyright © 2022 American Psychiatric Association

DSM, DSM-5, and DSM-5-TR are registered trademarks of the American Psychiatric Association. Use of these terms is prohibited without permission of the American Psychiatric Association.

ALL RIGHTS RESERVED. Unless authorized in writing by the APA, no part of this book may be reproduced or used in a manner inconsistent with the APA's copyright. This prohibition applies to unauthorized uses or reproductions in any form, including electronic applications.

Correspondence regarding copyright permissions should be directed to DSM Permissions, American Psychiatric Association Publishing, 800 Maine Avenue SW, Suite 900, Washington, DC 20024-2812.

Manufactured in the United States of America on acid-free paper.

ISBN 978-0-89042-575-6 (Hardcover) 1st printing February 2022

ISBN 978-0-89042-576-3 (Paperback) 1st printing February 2022

American Psychiatric Association
800 Maine Avenue SW
Suite 900
Washington, DC 20024-2812
www.psychiatry.org

The correct citation for this book is American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision. Washington, DC, American Psychiatric Association, 2022.

Library of Congress Cataloging-in-Publication Data

Names: American Psychiatric Association, issuing body.

Title: Diagnostic and statistical manual of mental disorders : DSM-5-TR / American Psychiatric Association.

Other titles: DSM-5-TR

Description: Fifth edition, text revision. | Washington, DC : American Psychiatric Association Publishing, [2022] | Includes index.

Identifiers: LCCN 2021051781 (print) | LCCN 2021051782 (ebook) | ISBN 9780890425756 (hardcover ; alk. paper) | ISBN 9780890425763 (paperback ; alk. paper) | ISBN 9780890425770 (ebook)

Subjects: MESH: Diagnostic and statistical manual of mental disorders. 5th ed | Mental Disorders—classification | Mental Disorders—diagnosis

Classification: LCC RC455.2.C4 (print) | LCC RC455.2.C4 (ebook) | NLM WM 15 | DDC 616.89/ 075—dc23/eng/20211209

LC record available at <https://lccn.loc.gov/2021051781>

LC ebook record available at <https://lccn.loc.gov/2021051782>.

British Library Cataloguing in Publication Data

A CIP record is available from the British Library.

Text Design—Tammy J. Cordova

Manufacturing—Sheridan Books, Inc.

Gender Dysphoria

In this chapter, there is one overarching diagnosis of gender dysphoria, with separate developmentally appropriate criteria sets for children and for adolescents and adults. The area of sex and gender is highly controversial and has led to a proliferation of terms whose meanings vary over time and within and between disciplines. An additional source of confusion is that in English “sex” connotes both male/female and sexuality. This chapter employs constructs and terms as they are widely used by clinicians from various disciplines with specialization in treating gender dysphoria. In this chapter, *sex* and *sexual* refer to the biological indicators of male and female (understood in the context of reproductive capacity), such as in sex chromosomes, gonads, sex hormones, and nonambiguous internal and external genitalia. Disorders of sex development or differences of sex development (DSDs) included the historical terms *hermaphroditism* and *pseudohermaphroditism*. DSDs include somatic intersex conditions such as congenital development of ambiguous genitalia (e.g., clitoromegaly, micropenis), congenital disjunction of internal and external sex anatomy (e.g., complete androgen insensitivity syndrome), incomplete development of sex anatomy (e.g., gonadal agenesis), sex chromosome anomalies (e.g., Turner syndrome; Klinefelter syndrome), or disorders of gonadal development (e.g., ovotestes).

Gender is used to denote the public, sociocultural (and usually legally recognized) lived role as boy or girl, man or woman, or other gender. Biological factors are seen as contributing, in interaction with social and psychological factors, to gender development. *Gender assignment* refers to the assignment as male or female. This occurs usually at birth based on phenotypic sex and, thereby, yields the *birth-assigned gender*, historically referred to as “biological sex” or, more recently, “natal gender.” *Birth-assigned sex* is often used interchangeably with birth-assigned gender. The terms *assigned sex* and *assigned gender* encompass birth-assigned sex/gender but also include gender/sex assignments and reassignments made after birth but during infancy or early childhood, usually in the case of intersex conditions. *Gender-atypical* refers to somatic features or behaviors that are not typical (in a statistical sense) of individuals with the same assigned gender in a given society and historical era; *gender-nonconforming*, *gender variant*, and *gender diverse* are alternative nondiagnostic terms. *Gender reassignment* denotes an official (and sometimes legal) change of gender. *Gender-affirming treatments* are medical procedures (hormones or surgeries or both) that aim to align an individual’s physical characteristics with their *experienced gender*. *Gender identity* is a category of social identity and refers to an individual’s identification as male, female, some category in between (i.e., *gender fluid*), or a category other than male or female (i.e., *gender neutral*). There has been a proliferation of gender identities in recent years. *Gender dysphoria* as a general descriptive term refers to the distress that may accompany the incongruence between one’s experienced or expressed gender and one’s assigned gender. However, it is more specifically defined when used as a diagnostic category. It does not refer to distress related to stigma, a distinct although

possibly co-occurring source of distress. *Transgender* refers to the broad spectrum of individuals whose gender identity is different from their birth-assigned gender. *Cisgender* describes individuals whose gender expression is congruent with their birth-assigned gender (also *nontransgender*). *Transsexual*, a historic term, denotes an individual who seeks, is undergoing,

512

or has undergone a social transition from male to female or female to male, which in many, but not all, cases also involves a somatic transition by gender-affirming hormone treatment and genital, breast, or other gender-affirming surgery (historically referred to as *sex reassignment surgery*).

Although not all individuals will experience distress from incongruence, many are distressed if the desired physical interventions using hormones and/or surgery are not available. The current term is more descriptive than the previous DSM-IV term *gender identity disorder* and focuses on dysphoria as the clinical problem, not identity per se.

Gender Dysphoria

Diagnostic Criteria

Gender Dysphoria in Children

F64.2

- A. A marked incongruence between one's experienced/expressed gender and assigned gender, of at least 6 months' duration, as manifested by at least six of the following (one of which must be Criterion A1):
1. A strong desire to be of the other gender or an insistence that one is the other gender (or some alternative gender different from one's assigned gender).
 2. In boys (assigned gender), a strong preference for cross-dressing or simulating female attire; or in girls (assigned gender), a strong preference for wearing only typical masculine clothing and a strong resistance to the wearing of typical feminine clothing.
 3. A strong preference for cross-gender roles in make-believe play or fantasy play.
 4. A strong preference for the toys, games, or activities stereotypically used or engaged in by the other gender.
 5. A strong preference for playmates of the other gender.
 6. In boys (assigned gender), a strong rejection of typically masculine toys, games, and activities and a strong avoidance of rough-and-tumble play; or in girls (assigned gender), a strong rejection of typically feminine toys, games, and activities.
 7. A strong dislike of one's sexual anatomy.
 8. A strong desire for the primary and/or secondary sex characteristics that

VIEWPOINT

The Medical Implications of Banning Transgender Youth From Sport Participation

Ellis Barrera, BA
Division of
Endocrinology, Boston
Children's Hospital,
Boston, Massachusetts.

Kate Millington, MD
Division of
Endocrinology, Boston
Children's Hospital,
Boston, Massachusetts;
and Department of
Pediatrics, Harvard
Medical School,
Boston, Massachusetts.

Jessica Kremen, MD
Division of
Endocrinology, Boston
Children's Hospital,
Boston, Massachusetts;
and Department of
Pediatrics, Harvard
Medical School,
Boston, Massachusetts.

Regular physical activity is essential for the health and well-being of children, adolescents, and adults. Participation in physical activity during childhood and adolescence has a positive impact on physical health throughout the life span. Moderate to vigorous aerobic exercise has been demonstrated to improve metabolic health in adolescents and to contribute to improved bone mineral density (BMD) during a period of critical growth. Exercise impacts mental health, leading to a decrease in depressive symptoms, with team sports participation specifically associated with lower rates of loneliness and social anxiety, effects that are more pronounced for those at greater psychologic risk.¹ Adolescence is a period of rapid social development and athletic participation can offer an important opportunity for developing social connections. The emotional, social, and physical benefits of exercise and sports participation are indisputable: physical activity during childhood and adolescence has lifelong positive impacts on both physical and mental health.

Transgender and gender-diverse youth (TGD) are particularly vulnerable to mental and physical comorbidities that are mitigated by physical activity. Thus, the benefits of exercise may be particularly significant for TGD youth. The past year has seen widespread legislative efforts to exclude TGD youth from organized sports, even though organized sports represent one of the most important opportunities for youth to engage in regular physical activity. Nine states have passed legislation banning TGD youth from participating in athletic teams concordant with their gender, and more bills are currently under consideration.² Notably, collegiate and professional athletic associations have not supported banning transgender athletes from participation in sports. In fact, the recent Olympic games featured several transgender athletes. Restricting sports participation threatens to worsen physical activity engagement among TGD youth, who already have lower rates of exercise than their cisgender peers.³

Fear of discrimination and violence are major barriers to physical activity participation for TGD individuals. Gender-segregated spaces, such as bathrooms, locker rooms, and sports teams are frequently cited by TGD individuals as feeling unsafe, with more than half reporting avoiding these spaces.⁴ Forcing TGD youth to use spaces discordant with their gender exacerbates perceptions of difference and will lead to further avoidance of physical activity for these children and adolescents. The health impact of such bans is likely to be significant. Specifically, these bans could have substantial impacts on the cardiovascular health, bone health, mental health, and cognitive development of TGD youth.

Restricting sports participation threatens to worsen existing risk factors for poor cardiovascular outcomes in

TGD youth. TGD adults have higher rates of cardiovascular events, which may represent a confluence of risk factors including the effects of gender-affirming hormones (GAH); minority stress; and lifestyle factors. Even prior to initiation of GAH, TGD youth have worse cardiovascular risk profiles than their peers.⁵ Increased access to physical activity for TGD youth is important for improving cardiovascular risk and mediating the expected changes that occur with GAH.

Restricting sports participation threatens to worsen existing risk factors for poor bone mineral density in TGD youth. TGD youth may be at increased risk of low BMD. For some TGD children and adolescents, use of gonadotropin-releasing hormone analogs/agonists (GnRHa), which limit sex steroid exposure during a period of peak bone accrual, may contribute to impaired bone mineralization. However, low BMD has been found to precede GnRHa treatment and is also found in TGD adults who never used GnRHa therapy.³ These findings suggest that there may be other risk factors for low bone density in TGD youth, such as limited engagement in weight-bearing exercise and reduced calcium intake. Legislation that limits access to sports for TGD youth has the potential to worsen exercise-dependent bone strength accrual by discouraging TGD youth from athletic participation at a critical moment in bone development.

Restricting sports participation has the potential to negatively impact neurocognitive development in TGD youth. Aerobic exercise leads to improved neural connectivity in adolescent brains, which may positively influence their executive functioning and cognitive abilities. Furthermore, adolescent participation in organized sports has been positively associated with academic performance, attention, planning, problem-solving, working memory, and inhibitory control.⁶ Adolescence is a time of important neurodevelopment, and restricting transgender youth from participating in youth sports during this critical time deprives them of the potential benefits of physical activity on neurodevelopment.

Restricting sports participation for TGD youth has the potential to impact the experiences of cisgender youth involved in school athletics. Several bills aimed at limiting the involvement of TGD youth in sports allow for genital inspections in cases where a child's gender is called into question by another coach or parent.⁷ The possibility of invasive examinations may decrease overall participation in athletics for both TGD and cisgender youth. Even in states without this clause, uneasiness with exclusionary sporting policies may decrease overall participation, whereas inclusionary policies have been associated with an increase in sports participation for all youth.

Beyond the TGD discrimination that bans of TGD youth participation in sport endorse,

Corresponding Author: Jessica Kremen, MD, Division of Endocrinology, Boston Children's Hospital, 300 Longwood Ave, Boston, MA 02115 (jessica.kremen@childrens.harvard.edu).

policies aimed at limiting or preventing physical activity will negatively impact the health of TGD children and adolescents. The wave of legislation seeking to roll back sports participation

for TGD youth is cause for concern for anyone invested in the health and well-being of all children, regardless of gender identity.

ARTICLE INFORMATION

Published Online: December 6, 2021.
doi:10.1001/jamapediatrics.2021.4597

Conflict of Interest Disclosures: Dr Millington reported grants from Doris Duke Charitable Foundation. No other disclosures were reported.

REFERENCES

1. Brière FN, Yale-Soulière G, Gonzalez-Sicilia D, et al. Prospective associations between sport participation and psychological adjustment in adolescents. *J Epidemiol Community Health*. 2018; 72(7):575-581. doi:10.1136/jech-2017-209656
2. Freedom for All Americans. Legislative Tracker: Anti-Transgender Legislation. <https://freedomforallamericans.org/legislative-tracker/>

[anti-transgender-legislation/](#). Accessed July 16, 2021.

3. Lee JY, Finlayson C, Olson-Kennedy J, et al. Low bone mineral density in early pubertal transgender/gender diverse youth: findings from the Trans Youth Care Study. *J Endocr Soc*. 2020;4(9):a065. doi:10.1210/endo/bvaa065
4. Kosciw JG, Greytak EA, Zongrone AD, Clark CM, Truong NL. The 2017 National School Climate Survey: The Experiences of Lesbian, Gay, Bisexual, Transgender, and Queer Youth in Our Nation's Schools. 2018. Accessed April 6, 2020. <https://www.glsen.org/research>.
5. Millington K, Finlayson C, Olson-Kennedy J, Garofalo R, Rosenthal SM, Chan Y-M. Association of high-density lipoprotein cholesterol with sex

steroid treatment in transgender and gender-diverse youth. *JAMA Pediatr*. 2021;175(5):520-521. doi:10.1001/jamapediatrics.2020.5620

6. Brooks SJ, Parks SM, Stamoulis C. Widespread positive direct and indirect effects of regular physical activity on the developing functional connectome in early adolescence. *Cereb Cortex*. 2021;31(10):4840-4852. doi:10.1093/cercor/bhab126
7. AP News. Florida House moves to limit transgender students in sports. <https://apnews.com/article/school-athletics-genetic-testing-bills-florida-bd489bfff68700e9cfc1674774b3d6b8>. Accessed July 26, 2021.

VIEWPOINT

The Lack of Helmets in Olympic Skateboarding—Skating Into Dangerous Territory

Mary O'Connor

Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success.

Setu Mehta, BA

Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success.

Ruth Milanaik, DO

Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success.

Corresponding

Author: Ruth Milanaik, DO, Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, 1983 Marcus Ave, Ste 130, Lake Success, NY 11042 (rmilanaik@northwell.edu).

Skateboarding made its debut as an Olympic sport at the 2020 Tokyo Games. Under current Olympic guidelines, skaters older than 18 years are not required to wear helmets during competition.¹ Accordingly, many competitors opted not to wear helmets, including all 3 medal winners in the men's street competition. This sets a dangerous example for young, impressionable fans of the sport, who may consider elite Olympic skaters to be role models.

Dangers of Skateboarding

In 2020, skateboard, scooter, and hoverboard injuries contributed to 217 646 emergency department visits. Six thousand of these injuries occurred in children younger than 4 years, 81 000 occurred in kids between the ages of 5 and 14 years, and 63 000 occurred in youth between the ages of 15 and 24 years.² The American Academy of Orthopedic Surgeons states that on average, almost half of all skateboarding injuries involve children (mostly boys) younger than 15 years. These injuries can range from cuts and bruises to sprains, strains, and fractures of the wrist, nose, jaw, and other bones. Severe injuries can include concussions and other head injuries. The American Academy of Pediatrics recommends that children younger than 5 years should never ride a skateboard, and children younger than 10 years should only skateboard with adult supervision. These recommendations are informed by children's particular vulnerability to these types of injuries; with a higher center of gravity, poorer balance, slower reaction times, and less coordination, children are less capable of controlling their falls. Children also lack the judgement to react rapidly to changing environmental conditions such as speed,

traffic, and other risks, putting them at high risk for injury while riding.³

Sadeghian et al⁴ examined emergency department data from skateboard-associated injuries from 2003 to 2012 among individuals younger than 18 years. Their analysis found that the majority of injured skateboarders had not been wearing helmets. Skateboarders were found to be 2.5 times more likely than those in their reference groups to become hospitalized. Other studies corroborate these alarming findings, with Tominaga et al⁵ reporting that of skateboarders who were admitted to the hospital, 75% presented with a head injury.

Recommendations for Safety While Skateboarding

According to the American Academy of Orthopaedic Surgeons, skateboarders should wear helmets, closed-toe and slip-resistant shoes, wrist guards, knee pads, elbow pads, and protective eye gear. Of these protective equipment requirements, obtaining the correct helmet is of the utmost importance to guard against serious head injuries. An adequate helmet should meet or exceed safety standards of the US Consumer Product Safety Commission, fit snugly on the head, and be replaced frequently.³

Some argue that while amateurs should wear helmets, Olympic skateboarders are above the ability level where helmets are needed, and it is thus unnecessary to hold them to the same standards as beginners. However, elite athletes are not immune to the hazards of skateboarding. Great Britain's youngest-ever summer Olympian, 11-year-old skateboarder Sky Brown, had a fall

Circulating Testosterone as the Hormonal Basis of Sex Differences in Athletic Performance

David J. Handelsman,^{1,2} Angelica L. Hirschberg,^{3,4} and Stephane Bermon^{5,6}

¹ANZAC Research Institute, University of Sydney, Concord, New South Wales 2139, Australia; ²Department of Andrology, Concord Hospital, Sydney, New South Wales 2139, Australia; ³Department of Women's and Children's Health, Karolinska Institutet, 171 76 Stockholm, Sweden; ⁴Department of Gynecology and Reproductive Medicine, Karolinska University Hospital, 171 76 Stockholm, Sweden; ⁵Laboratoire Motricité Humaine, Education, Sport, Santé, Université Côte d'Azur, 06000 Nice, France; and ⁶Health and Science Department, International Association of Athletics Federations, 98000 Monaco

ABSTRACT Elite athletic competitions have separate male and female events due to men's physical advantages in strength, speed, and endurance so that a protected female category with objective entry criteria is required. Prior to puberty, there is no sex difference in circulating testosterone concentrations or athletic performance, but from puberty onward a clear sex difference in athletic performance emerges as circulating testosterone concentrations rise in men because testes produce 30 times more testosterone than before puberty with circulating testosterone exceeding 15-fold that of women at any age. There is a wide sex difference in circulating testosterone concentrations and a reproducible dose-response relationship between circulating testosterone and muscle mass and strength as well as circulating hemoglobin in both men and women. These dichotomies largely account for the sex differences in muscle mass and strength and circulating hemoglobin levels that result in at least an 8% to 12% ergogenic advantage in men. Suppression of elevated circulating testosterone of hyperandrogenic athletes results in negative effects on performance, which are reversed when suppression ceases. Based on the nonoverlapping, bimodal distribution of circulating testosterone concentration (measured by liquid chromatography–mass spectrometry)—and making an allowance for women with mild hyperandrogenism, notably women with polycystic ovary syndrome (who are overrepresented in elite athletics)—the appropriate eligibility criterion for female athletic events should be a circulating testosterone of <5.0 nmol/L. This would include all women other than those with untreated hyperandrogenic disorders of sexual development and noncompliant male-to-female transgender as well as testosterone-treated female-to-male transgender or androgen dopers. (*Endocrine Reviews* 39: 803 – 829, 2018)

Virtually all elite sports are segregated into male and female competitions. The main justification is to allow women a chance to win, as women have major disadvantages against men who are, on average, taller, stronger, and faster and have greater endurance due to their larger, stronger muscles and bones as well as a higher circulating hemoglobin level. Hence, elite female competition forms a protected category with entry that must be restricted by an objective eligibility criterion related, by necessity, to the relevant sex-specific physical advantages. The practical need to establish an eligibility criterion for elite female athletic competition led the International Association of Athletic Federations (IAAF) to establish a rule in 2011, endorsed by the International Olympic Committee (IOC) in 2012, for hyperandrogenic women. That

IAAF regulation stated that for athletes to be eligible to compete in female events, the athlete must be legally recognized as a female and, unless she has complete androgen insensitivity, maintain serum testosterone <10 nmol/L. That IAAF eligibility rule was challenged by an athlete to the Court for Arbitration in Sports, which ruled in 2015 that, although an eligibility criterion was justified, there was insufficient evidence of the extent of the competitive advantage enjoyed by hyperandrogenic athletes who had circulating testosterone >10 nmol/L over female athletes with circulating testosterone in the normal female range. The Court for Arbitration in Sports suspended the rule pending receipt of such evidence. In that context, the present review presents the available evidence on the hormonal basis for the sex difference

ISSN Print: 0163-769X

ISSN Online: 1945-7189

Printed in USA

Copyright © 2018

Endocrine Society

This article has been published

under the terms of the

Creative Commons

Attribution License (CC BY;

<https://creativecommons.org/licenses/by/4.0/>).

Received: 28 January 2018

Accepted: 18 June 2018

First Published Online:

13 July 2018

ESSENTIAL POINTS

- It is widely accepted that elite athletic competitions should have separate male and female events
- The main justification is that men's physical advantages in strength, speed, and endurance mean that a protected female category, with objective entry criteria, is required
- Prior to puberty, there is no sex difference in circulating testosterone concentrations and athletic performance
- From male puberty onward, the sex difference in athletic performance emerges as circulating testosterone concentrations rise as the testes produce 30 times more testosterone than before puberty, resulting in men having 15- to 20-fold greater circulating testosterone than children or women at any age
- This wide, bimodal sex difference in circulating testosterone concentrations and the clear dose-response relationships between circulating testosterone and muscle mass and strength, as well as the hemoglobin level, largely account for the sex differences in athletic performance
- Based on the nonoverlapping, bimodal distribution of circulating testosterone concentration (measured by liquid chromatography–mass spectrometry) with 95% reference ranges of 7.7 to 29.4 nmol/L in healthy men and 0 to 1.7 nmol/L in healthy premenopausal women—making an allowance for women with the mild hyperandrogenism of polycystic ovary syndrome, who are overrepresented in elite athletics—the eligibility criterion for female athletic events should be a circulating testosterone concentration of <5.0 nmol/L

in athletic performance. It concludes that the evidence justifies a revised eligibility criterion of a threshold

circulating testosterone concentration of 5 nmol/L (measured by a mass spectrometry method).

Sex, Fairness, and Segregation in Sport

If sports are defined as the organized playing of competitive games according to rules (1), fixed rules are fundamental in representing the boundaries of fair sporting competition. Rule breaking, whether by breaching eligibility or competition rules, such as use of banned drugs, illegal equipment, or match fixing, creates unfair competitive advantages that violate fair play. Cheating constitutes a fraud against not just competitors but also spectators, sponsors, the sport, and the public. In the absence of genuine fair competition, elite sports would lose their wide popular appeal and ability to captivate and inspire with the authentic attraction of genuine contest between highly trained athletes.

Nevertheless, fairness is an elusive, subjective concept with malleable boundaries that may change over time as social concepts of fairness evolve. For example, until the late 19th century when organized sports trainers emerged, training itself was considered a breach of fairness because competition was envisaged at that time as a contest based solely on natural endowments. Similarly, sports once distinguished between amateurs and professionals. The concept of fairness has deep and complex philosophical roots mainly focused on notions of distributive justice. These considerations affect sports through the universal application of antidiscrimination and human rights legislation. Less attention is given to the philosophical basis of fair competition in elite sports, where the objectives are not egalitarian but aim to discover a hierarchy of achievement derived

from a mixture of unequal natural talent and individual training effort. Excellent, insightful discussion of the legal and moral complexities of sex and fair competition in elite sports from a legal scholar and former elite female athlete is available (2).

The terms *sex* and *gender* are often confused and used as if interchangeable. *Sex* is an objective, specific biological state, a term with distinct, fixed facets, notably genetic, chromosomal, gonadal, hormonal, and phenotypic (including genital) sex, each of which has a characteristic defined binary form. Whereas all facets of biological sex are almost always aligned so that assignment of sex at birth is straightforward, rare instances in which two or more facets of biological sex conflict constitute an intersex state, now referred to as disorders (or differences) of sex development (DSDs) (3). In contrast, *gender* is a subjective, malleable, self-identified social construct that defines a person's individual gender role and orientation. Prompted by biological, personal, and societal factors, volitional expression of gender can take on virtually any form limited only by the imagination, with some individuals asserting they have not just a single natal gender but two genders, none, a distinct third gender, or gender that varies (fluidly) from time to time. Hence, whereas gender is usually consistent with biological sex as assigned at birth, in a few it can differ during life. For example, if gender were the basis for eligibility for female sports, an athlete could conceivably be eligible to compete at the same Olympics in both female and male events. These features render the unassailable personal assertion of gender identity incapable of forming a fair, consistent sex classification in elite sports.

The strongest justification for sex classification in elite sports is that after puberty men produce 20 times more testosterone than women (4–7), resulting in circulating testosterone concentrations 15-fold higher than in children or women of any age. Age-grade competitive sporting records show no sex differences prior to puberty, whereas from the age of male puberty onward there is a strong and ongoing male advantage (8). The striking male postpubertal increase in circulating testosterone provides a major, ongoing, cumulative, and durable physical advantage in sporting contests by creating larger and stronger bones, greater muscle mass and strength, and higher circulating hemoglobin as well as possible psychological (behavioral) differences. In concert, these render women, on average, unable to compete effectively against men in power-based or endurance-based sports.

Sex classification in sports therefore requires proof of eligibility to compete in the protected (female) category. This deceptively simple requirement for fairness is taken for granted by peer female competitors who regard participation by males, or athletes with physical features closely resembling males, as unfair. This makes policing of eligibility inescapable for sports, to avoid unfair male participation in female events. However, such policing inevitably intrudes into highly personal matters so that it must be achieved with respect for dignity and privacy, demanding use of the least invasive, scientifically reliable means. Unsurprisingly, this dilemma has always been highly contentious since it first entered international elite sports in the early 20th century, and it has become increasingly prominent and contentious in recent decades; nevertheless, the requirement to maintain fair play in female events will not disappear as long as separate female competitions exist. During recent decades, there has been progressively better understanding of the complex biology of genetic sex determination and the impact of pubertal sexual maturation in establishing phenotypic sexual dichotomy in physical capabilities. These sex-dichotomous physical features form the basis of, but remain quite distinct from, adult gender roles and identity. During the last century, as knowledge grew, the attempts to formalize a scientific basis for the unavoidable necessity of policing eligibility for the female category have been continually challenged. Most recently, the increasing assertion of gender self-identification as a social criterion has further challenged the hegemony of biology for determining “sports sex,” Coleman’s apt term (2). Allowing subjective gender self-identification to become the sole criterion of sports sex would allow for gaming and perceptions of systematic unfairness to grow. The case for women’s sports being defined by sex rather than gender, including the consequences of acceding to gender-based classification, has been outlined (9) in arguing the importance of proper medical

management of athletes intending to compete in female events.

Separate male and female events in sports is a dominant form of classification that is superimposed on other graduated age group and weight classifications (e.g., in weightlifting, power lifting, wrestling, boxing, rowing), which reflect differences in strength, power, and speed to ensure fairness in terms of opportunity to win and, additionally, safety in contact sports. Age and weight classifications rely on objective criteria (birth date, weigh-in weight) for eligibility, and so should sex classification. Nevertheless, some power sports dependent on explosive strength and power (e.g., throwing events, sprinting) do not segregate weight classes, whereas other sports where height is an advantage (e.g., basketball, jockeys) do not have height classifications. These sports disproportionately attract athletes with greater weight and/or power-to-weight ratio or advantageous stature, respectively. If sex classification were eliminated, such open or mixed competitions would be dominated almost exclusively by men. It therefore seems highly unlikely that sex classification would ever be discarded, despite calls on philosophical or sociological grounds to end “gender” classification in sport (10).

Sex Difference in Circulating Testosterone Levels

Testosterone biosynthesis, secretion, and regulation in men and women

An androgen is a hormone capable of developing and maintaining masculine characteristics in reproductive tissues (notably the genital tract, as well as in other tissues and organs associated with secondary sexual characteristics and fertility) and contributing to the anabolic status of nonreproductive body tissues (11). The two dominant bioactive androgens circulating in mature mammals, including humans—testosterone and its more potent metabolite DHT—account for the development and maintenance of all androgen-dependent characteristics, and their circulating levels in men and nonpregnant women arise from steroids synthesized *de novo* in the testes, ovary, or adrenals (12).

The sexually undifferentiated gonads in the embryo develop into either ovaries or testes according to whether a Y chromosome (or at least the *sry* gene) is present. After birth and until puberty commences, circulating testosterone concentrations are essentially the same in boys and girls, other than briefly in the neonatal period of boys when higher levels prevail. The onset of male puberty, a brain-driven process triggered by a still mysterious hypothalamic or higher cerebral mechanism (13), initiates a hormonal cascade. In males, this leads to enhanced pituitary LH secretion that stimulates the 500 million Leydig cells in the testes

to secrete 3 to 10 mg (mean, 7 mg) of testosterone daily (4, 6, 7, 14, 15). This creates a very high local concentration of testosterone within the testis as well as a steep downhill concentration gradient into the bloodstream that maintains circulating testosterone levels at adult male levels, which are tightly regulated by strong negative hypothalamic feedback of circulating testosterone. In the absence of testes, these mechanisms do not function in females. In girls, serum testosterone increases during puberty (16), peaking at age 20 to 25 years before declining gradually with age (17, 18), but it remains <2 nmol/L at all ages, as determined by a reliable method (see below).

In adult women, circulating testosterone is derived from three roughly equal sources: direct secretion from the adrenal gland or the ovary and indirect extraglandular conversion (in liver, kidney, muscle, fat, skin) from testosterone precursors secreted by the adrenal and ovary. Only when circulating testosterone concentrations rise in male adolescents above the prepubertal concentrations does the virilization characteristic of men commence, progress, and endure throughout adult life, at least until old age (18). In combination, these different sources produce ~ 0.25 mg of testosterone daily so that throughout life women maintain circulating testosterone levels of <2 nmol/L. Circulating testosterone concentrations in women are subject to little dynamic physiological regulation. As a result, circulating testosterone concentrations in healthy premenopausal women are stable (nonfluctuating) and not subject to strong negative feedback by exogenous testosterone (as happens in men). Even the small rise (50%) at the time of the mid-cycle LH surge triggering ovulation (19) remains within the physiological range for premenopausal females.

Male and female reference ranges for circulating testosterone

A reliable threshold for circulating testosterone must be set using measurement by the reference method of liquid chromatography–mass spectrometry (LC-MS) rather than using one of the various available commercial testosterone immunoassays. The necessary reliance on steroid mass spectrometry for clinical applications in endocrinology, reproductive medicine, and sports medicine is widely recognized. It has been standard for decades in antidoping science (20), and the growing consensus is that it is required for high-quality clinical research and practice recognized by cognate professional societies (21, 22) and editorials in leading clinical endocrinology (23) and reproductive medicine (24) journals. The inherently limited specificity of testosterone immunoassays arises from antibody cross-reactivity with structurally related steroids (such as precursors and metabolites) other than the intended target. As a result, all steroid immunoassays, including for testosterone, display method-specific bias whereby, for example, the lower limit of a

testosterone reference range in healthy young men varies from 7.3 to 12.6 nmol/L according to the immunoassay used, so that no consensus definition of a lower limit could be obtained independent of the commercial immunoassay method used (25). Furthermore, testosterone immunoassays are optimized for circulating levels in men but display increasing inaccuracy at the lower, by an order of magnitude, circulating testosterone concentrations in women or children. In contrast to immunoassays, LC-MS–based methods are highly specific and do not depend on proprietary antibodies. Using LC-MS–based measurements, method-specific bias can be avoided and a fixed consensus lower reference limit defined (Table 1). Hence, for the precision required in sports medicine, whether for eligibility criteria or antidoping applications, testosterone in serum must be measured by LC-MS methods.

Prior to puberty, levels of circulating testosterone as determined by LC-MS are the same in boys and girls (16). They remain lower than 2 nmol/L in women of all ages. However, from the onset of male puberty the testes secrete 20 times more testosterone resulting in circulating testosterone levels that are 15 times greater in healthy young men than in age-similar women. Using LC-MS measurement, circulating testosterone in adults has a strikingly nonoverlapping bimodal distribution with wide and complete separation between men and women. Table 1 (25–36) summarizes data from appropriate reported studies using mass spectrometry–based methods to measure serum testosterone in healthy men and women. Based on a number-weighted pooling with conventional 95% two-sided confidence limits of the eight available studies using LC-MS measurements of serum testosterone, the reference range for healthy young men (18 to 40 years) is 7.7 nmol/L to 29.4 nmol/L. Similarly, summarizing the nine available studies for healthy menstruating women under 40 years, the 95% (two-sided) reference range is 0 to 1.7 nmol/L. These reference limits do not control for factors such as oral contraceptive use (35, 36), menstrual phase (19), SHBG (37, 38), overweight (39, 40), fasting and smoking (41), diet (40), and physical activity (42, 43) in women and men, all of which have small effects on circulating testosterone but without materially influencing the divergence between the nonoverlapping bimodal distribution of male and female reference ranges of circulating testosterone.

In creating a threshold for eligibility for female events it is also necessary to make allowance for women with polycystic ovary syndrome (PCOS) and nonclassical adrenal hyperplasia. PCOS is a relatively common disorder among women of reproductive ages with a prevalence of 6% to 10%, depending on the diagnostic criteria used (44), in which mild hyperandrogenism is a key clinical feature and has higher than expected prevalence among elite female athletes

Table 1. Serum Testosterone Measurements by LC-MS Methods in Studies of Healthy Men and Women

Study	Sample (Age 18–40 y)	N	Lower 95% CL (nmol/L)	Upper 95% CL (nmol/L)
Men				
Sikaris <i>et al.</i> , 2005 (25)	Elite, eugonadal	124	10.4	30.1
Turpeinen <i>et al.</i> , 2008 (26)	Convenience	30	10.1	31.2
Kushnir <i>et al.</i> , 2010 (27)	Convenience	132	7.2	24.2
Salameh <i>et al.</i> , 2010 (28)	Convenience	264	7.1	39.0
Neale <i>et al.</i> , 2013 (29)	Convenience	67	10.6	31.9
Kelsey <i>et al.</i> , 2014 (30)	Secondary pooled analysis	1058	7.2	25.3
Hart <i>et al.</i> , 2015 (31)	Birth cohort	423	7.4	28.0
Travison <i>et al.</i> , 2017 (32)	Pooled two cohorts	1656	7.9	31.1
Number-weighted mean			7.7	29.4
Women				
Turpeinen <i>et al.</i> , 2008 (26)	Convenience	32	0.8	2.8
Kushnir <i>et al.</i> , 2010 (27)	Convenience	104	0.3	2.0
Salameh <i>et al.</i> , 2010 (28)	Convenience	235	0.03	1.5
Haring <i>et al.</i> , 2012 (33)	Population-based	263	0.04	2.0
Neale <i>et al.</i> , 2013 (29)	Convenience	90	0	1.7
Bui <i>et al.</i> , 2013 (34)	Convenience	25	0.30	1.69
Rothman <i>et al.</i> , 2013 (19)	Convenience	31	0.4	0.92
Bermon and Garnier, 2017 (35)	Elite athletes	1652	0	1.62
Eklund <i>et al.</i> , 2017 (36)	Elite athletes and controls	223	0.26	1.73
Number-weighted mean			0.06	1.68

Abbreviation: CL, confidence limit.

(36, 45–47). Nonclassical adrenal hyperplasia is a milder and later (adult) onset variant of classical congenital adrenal hyperplasia (48) with a much higher but still rare population prevalence (1:1000 vs 1:16,000 for the classical variant) (49). Table 2 (50–64) summarizes clinical studies ($n = 16, \geq 40$ women) reporting serum testosterone concentrations measured by LC-MS in samples from women with PCOS.

The pooled data reveal that the upper limit of serum testosterone in women with PCOS is 3.1 nmol/L (95% CI, one-sided) or 4.8 nmol/L (using a 99.99% CI, one-sided) (Table 3). Hence, a conservative threshold for circulating testosterone of 5 nmol/L measured by LC-MS would identify <1:10,000 women with PCOS as false positives, based on circulating testosterone measurement alone. Circulating testosterone higher than this threshold is likely to be due to testosterone-secreting adrenal or ovarian tumors, intersex/DSD, badly controlled or noncompliant male-to-female (M2F) transgender athletes, or testosterone doping.

The physiological effects of testosterone depend on the circulating testosterone, not its source (endogenous or exogenous)

Testosterone, whether of a natural endogenous or manufactured exogenous source, has an identical chemical structure and biological effects, aside from minor differences in isotopic composition, which are biologically insignificant. At equivalent doses and circulating levels, exogenous testosterone exerts the same biological and clinical effects on every known androgen-responsive tissue or organ as endogenous testosterone, apart from effects on spermatogenesis, which as discussed below is only a matter of degree. Consequently, exogenous testosterone is a fully effective substitute for endogenous testosterone in therapeutic use, countering the effects of testosterone deficiency due to hypogonadism (reproductive system disorders). Any purported differences between endogenous and exogenous testosterone are due to corresponding differences in the endogenous production rate or exogenous dose. Such differences in

Data taken directly from paper or interpolated from other data (e.g., median, quartiles, ranges, sample size) supplied as described by Wan *et al.*, 2014 (Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 14: 135) are shown in italics.

Table 2. Summary of Serum Testosterone (nmol/L) by LC-MS in Women With PCOS From 16 Studies

Study	N	Mean	SD
Moran <i>et al.</i> , 2017 (50)	92	0.24	0.08
Münzker <i>et al.</i> , 2017 (51)	274	0.93	0.19
O'Reilly <i>et al.</i> , 2017 (52)	114	0.55	0.19
Handelsman <i>et al.</i> , 2017 (53)	152	0.38	0.25
Pasquali <i>et al.</i> , 2016 (54)	156	1.17	0.47
Yang <i>et al.</i> , 2016 (55)	1159	2.2	1.44
Tosi <i>et al.</i> , 2016 (56)	116	1.33	0.55
Daan <i>et al.</i> , 2015 (57)	170	1.64	0.53
Bui <i>et al.</i> , 2015 (58)	44	0.85	0.3
Keefe <i>et al.</i> , 2014 (59)	52	1.7	0.97
Yasmin <i>et al.</i> , 2013 (60)	165	1.99	1.02
Janse <i>et al.</i> , 2011 (61)	200	1.12	0.47
Jedel <i>et al.</i> , 2011 (62)	72	0.23	0.08
Legro <i>et al.</i> , 2010 (Mayo) (63)	596	2.12	0.89
Legro <i>et al.</i> , 2010 (Quest) (63)	596	1.98	0.97
Stener-Victorin <i>et al.</i> , 2010 (64)	74	1.53	0.62
Sum	4032		
Number-weighted mean		1.69	0.87

effective exposure lead to corresponding differences in circulating testosterone levels and its effects according to the dose-response curves for testosterone.

Similar to all hormones and drugs, over their effective range of biological activity the dose-response relationship for testosterone is usually a sigmoidal curve with lower and upper plateaus joined by a monotonically rising middle region, which may be linear in the natural scale but more often log-linear (linear on the log or similar transformed scale). In the middle portion of the typical sigmoidal dose-response curve for the same increase in testosterone dose (or concentration), the response would be increased in simple proportional (*i.e.*, linear) but more often on a logarithmic scale. In contrast, at the lower and upper plateaus of dose or concentrations, changes in testosterone exposure may evoke minimal or no response on the endpoint. For example, in women of any age circulating testosterone concentrations are along the lower plateau of the dose-response curve, so that increases in circulating testosterone concentrations within that lower plateau may have minimal or no effect. In female athletes with the mild hyperandrogenism of PCOS, higher performance has been shown (47), with their muscle mass and power performance correlating with androgen levels (36).

However, beyond these effects where endogenous testosterone concentrations are in the high-normal adult female range, it is only when the increases in circulating testosterone concentrations substantially and consistently exceed those prevailing in childhood (<2 nmol/L) and among women including those with PCOS (<5 nmol/L) that the effects would replicate the effects of rising testosterone concentrations of boys in middle to late puberty (typically >8 nmol/L), that is, the masculinizing effects of increased muscle, bone, and hemoglobin characteristics of men. As shown above, the circulating testosterone of most women never reaches consistently >5 nmol/L, a level that boys must sustain for some time to exhibit the masculinizing effects of male puberty.

In addition, the effects of testosterone are modulated in a form of fine tuning by the patterns of exposure, such as whether the circulating testosterone is delivered in the unphysiological steady-state format (*e.g.*, quasi-steady-state delivery by implant or transdermal products) or by the peak-and-trough delivery of injections, as opposed to the natural state of endogenous fluctuations in serum testosterone around the average adult male levels. However, these latter pattern effects are subtle and the dominant effect remains that of dose and average testosterone

concentrations in blood, however they arise. Furthermore, there is evidence that the androgen sensitivity of responsive tissues differs and may be optimal at different circulating testosterone concentrations (65).

Male sexual function is maintained by endogenous testosterone at adult male circulating concentrations. These effects can be replicated by exogenous testosterone if and only if it achieves comparable circulating testosterone concentrations. For example, in a well-controlled prospective study of older men with prostate cancer (66), androgen deprivation achieving castrate levels of circulating testosterone sustained during 12 months markedly suppressed sexual desire and function, whereas those effects did not occur in age-matched men having nonhormonal treatment of prostate cancer or those without prostate cancer. In healthy younger men whose endogenous testosterone was fully suppressed, sexual function completely recovered when circulating testosterone was restored to the physiological male range by administration of exogenous testosterone (67). Similar effects were also observed in healthy, middle-aged men in whom male sexual function was fully maintained (compared with placebo) during 2 years of treatment with an exogenous androgen (DHT) despite that treatment causing sustained, complete suppression of endogenous testosterone (68). This further supports the key interpretation that the biological effects of exogenous or endogenous testosterone are the same at comparable circulating levels.

Clinically, exogenous testosterone replicates fully all effects of endogenous testosterone on every reproductive and nonreproductive organ or tissue, with the sole exception of the testis. Sperm production in the testis requires a very high concentration of testosterone (typically 100-fold greater than in the general bloodstream), which is produced in nature only by the action of the pituitary hormone LH. LH stimulates the Leydig cells in the interstitial space of the testis between seminiferous tubules to produce high intratesticular concentrations of testosterone, which are necessary and sufficient to initiate and maintain sperm production in the adjacent seminiferous tubules. This

high concentration of testosterone also provides a downhill gradient to supply the rest of the body, where circulating testosterone acts on androgen-responsive tissues to produce and maintain masculine patterns of androgenization. When exogenous testosterone (or any other androgen) is administered to men, pituitary LH is suppressed by negative feedback and the sperm production halts for as long as exogenous testosterone or androgen exposure continues, after which it recovers (69). However, even the reduction in spermatogenesis and testis size when men are treated with exogenous testosterone is only a matter of degree. It is well established in rodents (70, 71) that spermatogenesis is induced by exogenous testosterone when the testosterone concentrations in the testis are high enough to replicate what occurs naturally via LH stimulation (72). However, direct replication that high-dose testosterone also initiates and maintains spermatogenesis in humans is not feasible, as these testosterone doses are 10- to 100-fold higher than could be safely given to humans. Nevertheless, confirmatory evidence in humans is available from rare cases of men with an activating mutation of the chorionic gonadotropin/LH receptor (73, 74). This mutation causes autonomous testicular testosterone secretion leading to precocious puberty arising from the premature adult male circulating testosterone concentrations that lead to complete suppression of circulating gonadotropin (LH, FSH) secretion. In this illustrative case the testis was exposed to non-physiologically high testosterone concentrations (but without any gonadotropin stimulation) that induced sperm production and allowed for natural paternity (73). This indicates that even for spermatogenesis, exogenous testosterone can replicate all biological effects of endogenous testosterone in accordance with the relevant dose-response characteristics.

The most realistic view is that increasing circulating testosterone from the childhood or female range to the adult male range will have the same physiological effects whether the source of the additional testosterone is endogenous or exogenous. This is strongly supported by well-established knowledge about the relationship of circulating testosterone concentrations

Table 3. Upper Confidence Limits on Serum Testosterone in Women With PCOS

Confidence Interval	Likelihood ^a	SD ^b	One-Sided ^c	Two-Sided ^c
95%	1:20	1.96	3.13	3.39
99%	1:100	2.35	3.47	3.73
99.9%	1:1000	3.10	4.21	4.39
99.99%	1:10,000	3.72	4.77	4.95

^aLikelihood that a woman with PCOS would exceed that limit by chance.

^bNumber of SDs for each confidence limit.

^cTwo-sided CIs are conventional for a result that could exceed or fall below confidence limits, but here we focus only on values exceeding the upper limit, so that one-sided confidence limits are appropriate.

with the timing and manifestations of male puberty. The characteristic clinical features of masculinization (e.g., muscle growth, increased height, increased hemoglobin, body hair distribution, voice change) appear only if and when circulating testosterone concentrations rise into the range of males at mid-puberty, which are higher than in women at any age even after the rise in circulating testosterone in female puberty. If and only if the pubertal rise in circulating testosterone fails will the males affected be clinically considered hypogonadal. Such a failure of male puberty may occur for genetic reasons (arising from mutations that inactivate any of the cascade of proteins whose activity is critical in the hypothalamus to trigger male puberty) or as a result of acquired conditions, caused by pathological disorders of the hypothalamus or pituitary or functional defects arising from severe deficits of energy or nutrition (e.g., extreme overtraining, undernutrition), with the latter being comparable with hypothalamic amenorrhea or anorexia nervosa in female athletes/ballet dancers. If male puberty fails, testosterone replacement therapy is fully effective in replicating all of the distinctive masculine features apart from spermatogenesis.

Elevated circulating testosterone concentration caused by DSDs

Rare genetic intersex conditions known as DSDs can lead to markedly increased circulating testosterone in women. When coupled with ambiguous genitalia at birth, they may appear as undervirilized males or virilized females. This can cause athletes who were raised and identify as women to have circulating testosterone levels comparable to those of men and greatly exceeding those of non-DSD (and nondoped) women, including those with PCOS. Key congenital disorders in this category are 46,XY DSDs, namely 5 α reductase deficiency (75), 17 β -hydroxysteroid dehydrogenase type 3 deficiency (76), and androgen insensitivity (77, 78), as well as congenital adrenal hyperplasia (79), which is a 46,XX DSD. There is evidence that the first three conditions, components of 46,XY DSDs, are 140-fold more prevalent among elite female athletes than expected in the general population (80).

Genetic 5 α reductase deficiency is due to an inactivating mutation in the 5 α reductase type II enzyme (75). This leads to a deficit of DHT during fetal life when DHT is required for converting the sex-undifferentiated embryonic and fetal tissue to form the sex-differentiated masculine form external genitalia. Although genetic males (46,XY) with 5 α reductase deficiency will develop testes, they usually remain undescended and labial fusion to form a scrotum and phallic growth does not occur. Hence, at birth the external genitalia may appear feminine, leading to a female assigned natal sex. Thus, individuals with 5 α reductase deficiency may have male chromosomal sex

(46,XY), gonadal sex (testes), and hormonal sex (adult male testosterone concentrations), but such severely undervirilized genitalia that affected individuals may be raised from birth as females rather than as undervirilized males. However, from the onset of male puberty, testicular Leydig cells start producing large amounts of testosterone, and the steep rise in circulating testosterone to adult male levels (with the permissive role of 5 α reductase activity) leads to masculine virilization, including male patterns of muscle and bone growth, hemoglobin levels, and other masculine body habitus features (hair growth pattern, voice change), as well as phallic growth (80). Such changes of male puberty prompt around half affected individuals who had female sex assigned at birth and developed as girls prior to puberty to adopt a male gender identity and role at puberty (81). Sperm are formed in the testes so that, using *in vitro* fertilization, these individuals may father children (82).

17 β -Hydroxysteroid dehydrogenase type 3 deficiency (76) has a natural history similar to that of 5 α reductase deficiency. This disorder is due to inactivating mutations in a steroidogenic enzyme expressed only in the testis and that is essential for testosterone formation in the fetus. In the absence of a functional enzyme, the testis makes little testosterone but instead secretes large amounts of androstenedione, the steroid immediately prior to the enzymatic block. In the circulation, the excess of androstenedione is converted to testosterone (mainly by the enzyme AKR1C3) (12). Although the circulating testosterone is then converted to circulating DHT, insufficient DHT is formed locally within the urogenital sinus to virilize genitalia at birth. This causes the same severe undervirilization of the external genitalia of genetically male individuals, leading to ambiguous genitalia at birth despite male chromosomal, gonadal, and hormonal sex. When puberty arrives, the testes start producing the adult male testosterone output. Again, this leads to marked virilization and subsequent assumption of a male gender identity by some affected individuals, conflicting with a female assigned natal sex and childhood upbringing.

Androgen insensitivity, which arises from mutation in the androgen receptor (AR), poses different but complex challenges for eligibility for female athletic events. As the AR is located on the X chromosome, genetic males (46,XY) are hemizygous, so that an inactivating mutation in the AR can be partially or fully insensitive to androgen action. Affected individuals have male internal genitalia (testes in the inguinal canal or abdomen with Wolffian ducts) and consequently adult male circulating testosterone concentrations after puberty. These nonlethal mutations have a wide spectrum of functional effects, ranging from full resistance to all androgen action in complete androgen insensitivity syndrome (CAIS) where individuals have a full female phenotype with

normal female external genitalia, to partial androgen insensitivity syndrome (PAIS) where some androgen action is still exerted, leading to various degrees of ambiguous genitalia, or to mild androgen insensitivity, which produces a very mild, undervirilized male phenotype (normal male genital and somatic development but with little body hair and no male pattern balding) (77). Testosterone (and dihydrotestosterone) have no consistent effect of inducing normal nitrogen retention (anabolic) responses in patients with CAIS (83–86), although some reduced androgen responsiveness is retained by patients with PAIS (84, 87–90). Athletes with CAIS can compete fairly as females because the circulating testosterone, although at adult male levels, has no physiological effect so that, in terms of androgen action and the ensuing physical somatic advantages of male sex, affected individuals are indistinguishable from females and gain no benefits of the sex difference arising from unimpeded testosterone action. A more complex issue arises with athletes having PAIS reflecting the degree of incomplete impairment of AR function. Residual androgen action in such AR mutations is harder to characterize quantitatively, as there is no standardized, objective *in vitro* test to quantify AR functionality. Hence, individuals with PAIS may have adult male circulating testosterone concentrations but variable androgen sensitivity. At present, determination of eligibility to compete in the female category requires a case-by-case evaluation, primarily based on the degree of virilization. The current best available clinical approach to determining the functional impact (degree of functionality/sensitivity) of an AR mutation is based on the degree of somatic, primarily genital, virilization assessed according to the Quigley classification of grade of androgen sensitivity (91).

Congenital adrenal hyperplasia (CAH) is a relatively common defect in adrenal steroidogenesis in the enzymatic pathway, leading to synthesis of cortisol, aldosterone, and sex steroid precursors. The disease varies in severity from life-threatening (adrenal failure) to mild (hirsutism and menstrual irregularity), or even asymptomatic and undiagnosed. The most common mutations causing CAH occur in the 21-hydroxylase enzyme, accounting for 95% of cases (79). The defect leads to a bottleneck, creating a major backing up of precursor steroids that then overflow into other steroid pathways, leading to diagnostic high levels of 17-hydroxyprogesterone and, in female patients, excessive circulating testosterone or other adrenal-source androgen precursors (e.g., androstenedione, dehydroepiandrosterone) that may be converted to testosterone in tissues. A common clinical problem with management of CAH is that glucocorticoid/mineralocorticoid treatment is not always fully effective partly due to variable compliance, which may leave high circulating testosterone, including well into or even above the normal male range (92). It is unlikely

that mild nonclassical congenital adrenal hyperplasia is a major contributor to the mild hyperandrogenism prevalent among elite female athletes. The prevalence of PCOS (6% to 16%) is about 100-fold higher than mild nonclassical congenital adrenal hyperplasia (0.1%) (49), whereas a disproportionately high number of elite female athletes (especially in power sports) have PCOS (45). In one study of hyperandrogenic female athletes, even mild nonclassical adrenal hyperplasia was ruled out by normal 17-hydroxyprogesterone (36) and, in another (47), reported serum androstenedione and cortisol did not differ from controls, ruling out significant congenital adrenal hyperplasia.

Sex Difference in Muscle, Hemoglobin, Bone, and Athletic Performance Relating to Adult Circulating Testosterone Concentrations

Following puberty, testosterone production increases (16) but remains <2 nmol/L in women, whereas in men testosterone production increases 20-fold (from 0.3 mg/d to 7 mg/d), leading to 15-fold higher circulating testosterone concentrations (15 vs 1 nmol/L). The greater magnitude of sex difference in testosterone production (20-fold) compared with circulating levels (15-fold) is due to women's higher circulating SHBG, which retards testosterone clearance, creating a slower circulating half-time of testosterone. This order-of-magnitude difference in circulating testosterone concentrations is the key factor in the sex difference in athletic performance due to androgen effects principally on muscle, bone, and hemoglobin.

Muscle

Biology

It has been known since ancient times that castration influences muscle function. Modern knowledge of the molecular and cellular basis for androgen effects on skeletal muscle involves effects due to androgen (testosterone, DHT) binding to the AR that then releases chaperone proteins, dimerizes, and translocates into the nucleus to bind to androgen response elements in the promoter DNA of androgen-sensitive genes. This leads to increases in (1) muscle fiber numbers and size, (2) muscle satellite cell numbers, (3) numbers of myonuclei, and (4) size of motor neurons (93). Additionally, there is experimental evidence that testosterone increases skeletal muscle myostatin expression (94), mitochondrial biogenesis (95), myoglobin expression (96), and IGF-1 content (97), which may augment energetic and power generation of skeletal muscular activity.

Customized genetic mouse models can provide unique experimental insight into mammalian physiology that is unobtainable by human experimentation.

"Sex differences in height, where they exist, are largely dependent on postpubertal differences in circulating testosterone."

The tight evolutionary conservation of the mammalian reproductive system explains why genetic mouse models have provided consistent, high-fidelity replication of the human reproductive system (98, 99). Genetic males (46,XY) with androgen insensitivity displaying similar features occur through the spontaneous production of inactivating AR mutations in all mammalian species studied, including humans, where they are known as women with CAIS. The converse, genetic females (46,XX) resistant to all androgen action cannot occur naturally in humans or other mammals. This is because fully androgen-resistant females must have both X chromosomes carrying an inactivated AR. In turn, this requires acquiring one X chromosome from their father, and hemizygous males bearing a single X chromosome with an inactive AR produce no sperm, as a functional AR is biologically indispensable for making sperm in any mammal. However, androgen-resistant females can be bred by genetic engineering using the Cre-Lox system (100). An important finding from such studies is that androgen-resistant female mice have essentially the same muscle mass and function as wild-type androgen-sensitive females bearing normal AR, whereas androgen-resistant male mice have smaller and weaker muscle mass and function than do wild-type males and comparable instead with wild-type females (101). This indicates that androgen action, represented by circulating testosterone, is the key determinant of the higher muscle mass and strength characteristic of males compared with females. Furthermore, endogenous circulating testosterone has minimal effects on skeletal muscle mass and strength in female mice because of its low levels. Although these experiments cannot be replicated in humans, their key insight is that the higher circulating testosterone in males is the determinant of the male's greater muscle mass and function compared with females. Nevertheless, there is also evidence that hyperandrogenic women, mostly with PCOS, have increased muscle mass and strength that correlates with mildly increased circulating testosterone in the high-normal female range (36, 47).

Observational data

There is a clear sex difference in both muscle mass and strength (102–104) even adjusting for sex differences in height and weight (104, 105). On average, women have 50% to 60% of men's upper arm muscle cross-sectional area and 65% to 70% of men's thigh muscle cross-sectional area, and women have 50% to 60% of men's upper limb strength and 60% to 80% of men's leg strength (106). Young men have on average a skeletal muscle mass of >12 kg greater than age-matched women at any given body weight (104, 105). Whereas numerous genes and environmental factors (including genetics, physical activity, and diet) may contribute to muscle mass, the major cause of the sex

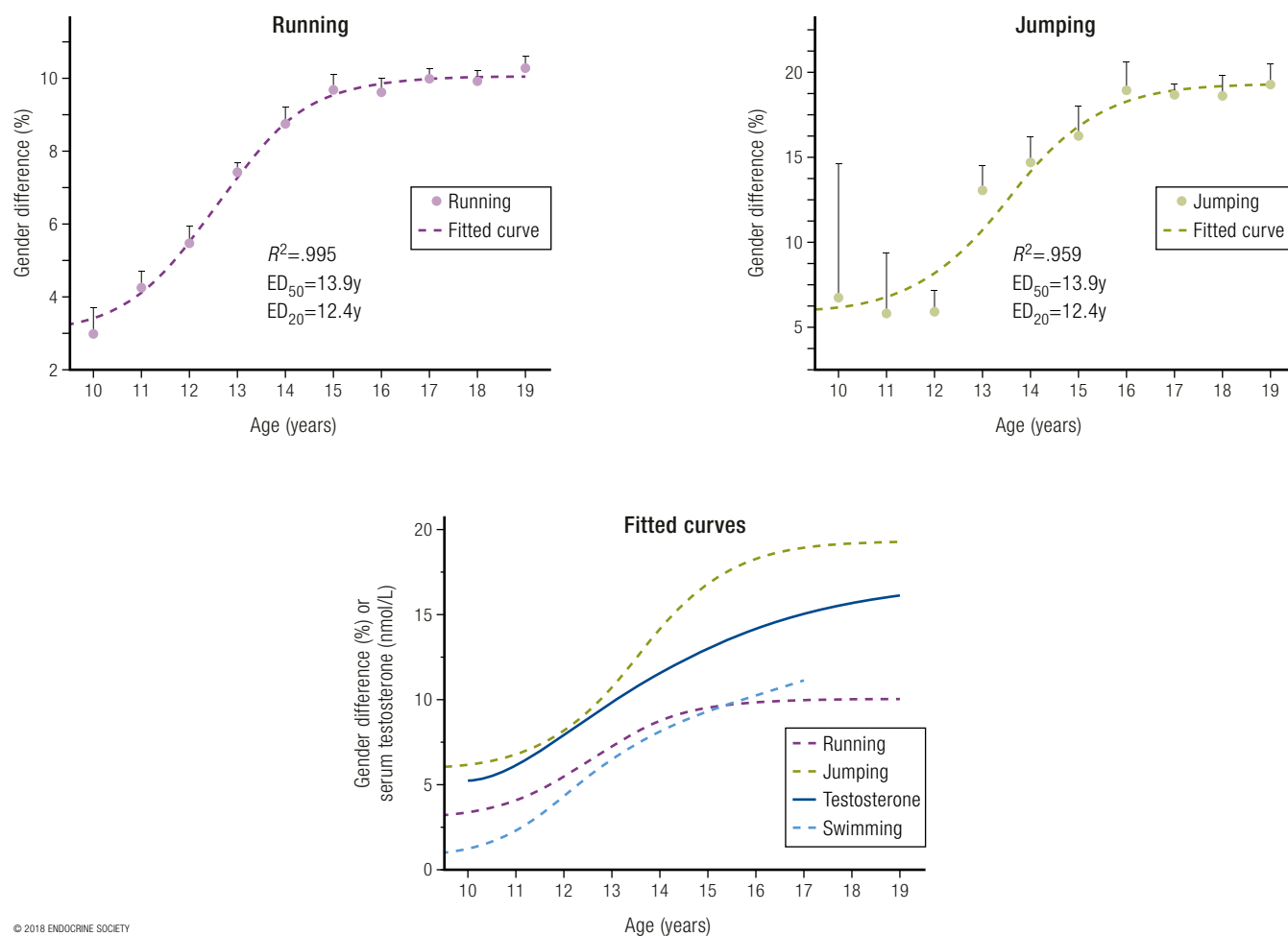
difference in muscle mass and strength is the sex difference in circulating testosterone.

Age-grade competitive sports records show minimal or no female disadvantage prior to puberty, whereas from the age of male puberty onwards there is a strong and ongoing male advantage. Corresponding to the endogenous circulating testosterone increasing in males after puberty to 15 to 20 nmol/L (sharply diverging from the circulating levels that remain <2 nmol/L in females), male athletic performances go from being equal on average to those of age-matched females to 10% to 12% better in running and swimming events, and 20% better in jumping events (8) (Fig. 1). Corroborative findings are provided by a Norwegian study that examined performance of adolescents in certain athletic events but without reference to contemporaneous circulating testosterone concentrations (107). The striking postpubertal increase in male circulating testosterone provides a major, ongoing, cumulative, and durable advantage in sporting contests by creating greater muscle mass and strength. These sex differences render women unable to compete effectively against men, especially (but not only) in power sports.

These findings are supported by studies of non-athletic women showing that muscle mass is increased in proportion to circulating testosterone in women with mildly elevated testosterone levels due to PCOS (108, 109), a condition that is more prevalent among elite female athletes who exhibit these features (36, 45, 47), often undiagnosed (46), but that may provide an ergogenic advantage (47), consistent with the graded effects of circulating testosterone on explosive performance in men and women (110).

Studies of elite female athletes further corroborate these findings. One study demonstrates dose-response effects of better performance in some (400 m running, 400 m hurdles, 800 m running, hammer throw, pole vault) but not all athletic events correlated with significantly higher endogenous testosterone in female, but not male, athletes. Even within the low circulating testosterone levels prevailing within the normal female range, in these events there was a significant advantage of 1.8% to 4.5% among those in the highest tertile compared with the lowest tertile of endogenous testosterone (35). A further study of elite female athletes corroborates and extends these observations in that endogenous androgens are associated with a more anabolic body composition as well as enhanced muscular performance (36). In this study, 106 Swedish Olympic female athletes were compared with 117 age- and weight (body mass index)-matched sedentary control women for their muscle and bone mass (by dual-energy X-ray absorptiometry), their muscular strength (squat and countermovement jumps), and testosterone and DHT, as well as androgen precursors (dehydroepiandrosterone, androstenedione) and urinary androgen glucuronide metabolites (androsterone,

Figure 1. Sex differences in performance (in percentage) according to age (in years) in running events, including 50 m to 2 miles (upper left panel), and in jumping events, including high jump, pole vault, triple jump, long jump, and standing long jump (upper right panel) [for details, see Ref. (8)]. The lower panel is a fitted sigmoidal curve plot of sex differences in performance (in percentage) according to age (in years) in running, jumping, and swimming events, as well as the rising serum testosterone concentrations from a large dataset of serum testosterone of males. Note that in the same dataset, female serum testosterone concentrations did not change over those ages, remaining the same as in prepubertal boys and girls. Data are shown as mean and SEM of the pooled sex differences by age. Reproduced with permission from Handelsman DJ. Sex differences in athletic performance emerge coinciding with the onset of male puberty. *Clin Endocrinol (Oxf)*. 2017;**87**:68–72.



© 2018 ENDOCRINE SOCIETY

etiocholanolone, 3 and 17 3α -diols) measured by LC-MS (36). The athletes displayed higher muscle (and bone) mass than did the sedentary control women, with strength tests correlating strongly with muscle mass whether in total or just in the legs. In turn, muscle mass and strength were correlated with androgens and androgen precursors. Considering that such studies may be confounded by factors such as menstrual phase and dysfunction, as well as heterogeneous sports disciplines, which weaken the power of the study, these findings can be regarded as quite robust.

Interventional data

Dose-response studies show that in men whose endogenous testosterone is fully suppressed, add-back administration of increasing doses of testosterone that produce graded increases in circulating testosterone causes a

dose-dependent (whether expressed according to testosterone dose or circulating levels) increase in muscle mass (measured as lean body mass) and strength (65, 111). Taken together, these studies prove that testosterone doses leading to circulating concentrations from well below to well above the normal male range have unequivocal dose-dependent effects on muscle mass and strength. These data strongly and consistently suggest that the sex difference in lean body mass (muscle) is largely, if not exclusively, due to the differences in circulating testosterone between men and women. These findings have strong implications for power-dependent sport performance and largely explain the potent efficacy of androgen doping in sports.

The key findings providing conclusive evidence that testosterone has prominent dose-response effects in men are reported in studies by Bhasin and colleagues that proved a monotonic dose response,

extending from subphysiological to supraphysiological range for men for testosterone effects on muscle mass, size, and strength in healthy young men, findings that have been replicated and confirmed by an independent group (65). Both sets of studies used a common design of fully suppressing all endogenous testosterone (to castrate levels) for the full duration of the experiment by administering a GnRH analog. In the Bhasin and colleagues studies, participants were then randomized to five groups and each received weekly injections of 25 mg, 50 mg, 125 mg, 300 mg, or 600 mg of testosterone enanthate for 20 weeks. In effect, this was two subphysiological and two supraphysiological testosterone doses. In these studies, the lowest testosterone dose produced a mean serum testosterone of 253 ng/dL (8.8 nmol/L) in younger men and 176 ng/dL (6.1 nmol/L) in older men. The studies showed a consistent dose response for muscle mass and strength that was clearly related to testosterone dose and consequential blood testosterone concentrations (Fig. 2, upper panel).

The study of Finkelstein *et al.* (65) involved the same design and involved 400 healthy men aged 20 to 50 years who had complete suppression of endogenous testosterone for the 16 weeks of the study, with testosterone added back using daily doses of 0, 1.25 g, 2.5 g, 5 g, or 10 g of a topical 1% testosterone gel. This again created a graded dose-response curve for serum testosterone and for muscle mass and strength. The inclusion of a 0 (placebo) dose allowed differentiation between the 0 and lowest testosterone dose. The placebo (0) dose produced a serum testosterone of 0.7 nmol/L (the typical mean for castrated men, childhood, and women of any age). Meanwhile, the lowest testosterone dose (1.25 g of gel per day) produced a serum testosterone of 6.9 nmol/L, which is equivalent to that of a male in early to middle puberty. A key finding for this review is that, from this study of men, the increase in serum testosterone from mean of normal female concentration (0.9 nmol/L) to supraphysiological female concentrations (6.9 nmol/L) produced significant increases of 2.3% for total body lean (muscle) mass, 3.0% for thigh muscle area, and 5.5% increase in leg press strength (digitized data pooling of both cohorts from lower panel, Fig. 2).

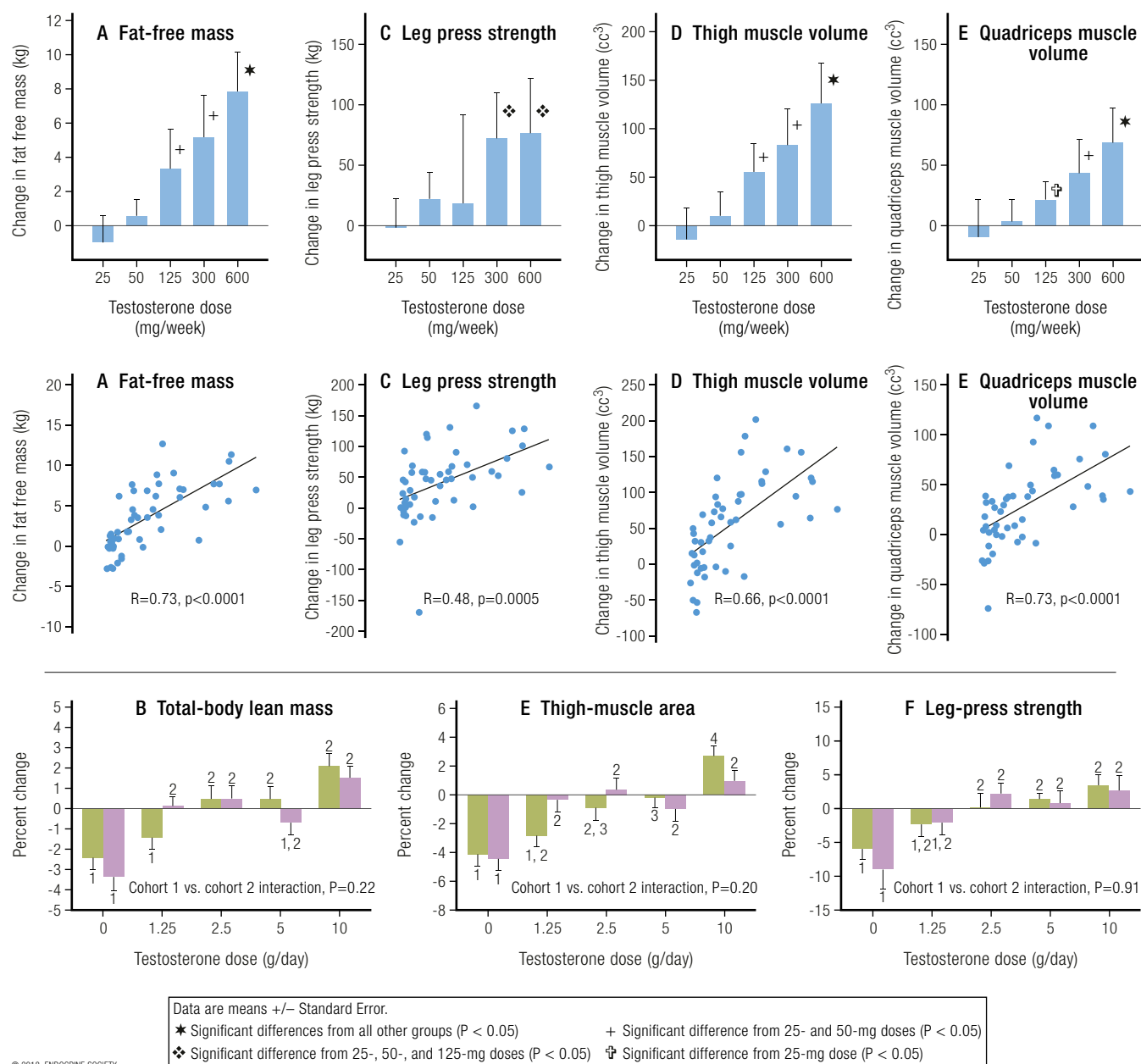
Studies of the ergogenic effects of supraphysiological concentrations of circulating testosterone require studies administering graded doses of exogenous testosterone for months. Owing to ethical concerns regarding risks of unwanted virilization and hormone-dependent cancers, however, few studies have administered supraphysiological testosterone doses to healthy women. One well-designed, randomized placebo-controlled study of postmenopausal women investigated the effects of different testosterone doses on muscle mass and performance and physical function (112). Sixty-two women (mean age, 53 years) all had a standard estrogen-replacement dose administered during a 12-week run-in period (to

eliminate any hypothetical confounding effects of estrogen deficiency), after which they were randomized to one of five groups receiving weekly injections of testosterone enanthate (doses: 0, 3 mg, 6.25 mg, 12.5 mg, and 25 mg, respectively) for 24 weeks. The increasing doses of testosterone produced an expected dose response in serum testosterone concentrations (by LC-MS), with the highest testosterone dose (25 mg/wk) producing a mean nadir concentration of 7.3 nmol/L. The women whose testosterone concentrations were increased to 7.3 nmol/L achieved significant increases in muscle mass and strength (Table 4), ranging from 4.4% for muscle (lean) mass to between 12% and 26% for measures of muscle strength (chest and leg press, loaded stair climb). As muscle strength measurement is effort-dependent, the placebo-controlled design of the Huang *et al.* (112) study supports the further interpretation that the highest dose of testosterone also had prominent mental motivational effects in the effort-dependent tests of muscle strength. These findings provide salient direct evidence of the ergogenic effects of hyperandrogenism in female athletes confirming that at least up to average circulating testosterone concentrations of 7.3 nmol/L, women display a dose-response relationship similar to that of men, with supraphysiological doses of testosterone leading to significant gains in muscle mass and power.

These effects of testosterone administration on circulating testosterone concentrations and muscle mass and strength in females may be compared with the effects in males from the Finkelstein *et al.* (65) and Bhasin and colleagues studies. In men, the lowest testosterone dose (1.25 g/d) increased mean serum testosterone to 6.9 nmol/L (equivalent to levels seen in early to middle male puberty), resulting in significant increases of total body lean (muscle) mass (2.3%), thigh muscle area (3.0%), and leg press strength (5.5%) compared with the placebo dose that resulted in a serum testosterone of 0.7 nmol/L. In the Huang *et al.* (112) study (Fig. 3), muscle mass and strength in postmenopausal women displayed a flat response at the three lower doses, when circulating testosterone concentrations remain <5 nmol/L, and displayed a significant increase only when the mean circulating testosterone concentration produced by the highest testosterone dose first increased circulating testosterone concentrations >5 nmol/L. This pattern, flat at lower doses and rising at the highest dose, represents the lower plateau and the earliest rising portion, respectively, of the sigmoidal dose-response curve of testosterone for muscle.

Data corroborating the Huang *et al.* study results comes from another well-controlled study in which postmenopausal women who were administered methyl testosterone following a run-in period of estrogen replacement displayed a significant increase in lean (muscle) mass as well as upper and lower limb

Figure 2. Strong dose-response relationship between testosterone dose and circulating concentration with muscle mass and strength in men. The upper panels [from Bhasin *et al.* (111)] display the strong dose-response relationships of muscle mass shown as (A) “lean” or “fat-free” mass or volume of (D) thigh and (E) quadriceps muscle and (C) of leg muscle strength with increasing testosterone dose (upper row) or circulating concentration (middle row). Serum testosterone concentrations are in US units (ng/dL; divide by 28.8 to get nmol/L). Adapted with permission from Bhasin S, Woodhouse L, Casaburi R, *et al.* Testosterone dose-response relationships in healthy young men. *Am J Physiol Endocrinol Metab.* 2001;281:E1172–E1181. The lower panels [from Finkelstein *et al.* (65)] show the strong dose-response relationships of (B) whole-body muscle mass, (E) thigh muscle mass, and (F) leg press strength with increasing testosterone dose. Cohorts 1 and 2 were treated with the same increasing doses of testosterone but either without (green fill, cohort 1) or with (purple fill, cohort 2) an aromatase inhibitor (anastrozole), which prevents conversion of testosterone to estradiol. The differences between cohorts (*i.e.*, use of anastrozole) was not significant for muscle mass and strength and can be ignored with results of the two cohorts being pooled. Reproduced with permission from Finkelstein JS, Lee H, Burnett-Bowie SA, Pallais JC, *et al.* Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med* 2013;369:1011–1022.



power during a 16-week double-blind, parallel group study (113).

Similarly, two prospective studies of the first 12 months of treatment of transmen [female-to-male

(F2M) transgender] shows a consistent major increase in muscle mass and strength due to testosterone administration. In one study testosterone treatment of 17 transmen achieving adult male circulating testosterone levels

Table 4. Effects of Testosterone on Muscle Mass and Strength in Women

Androgen-Sensitive Variable	Baseline	Increase	% Increase
Lean muscle mass, kg	43 ± 6	1.9 ± 0.5	4.4
Chest press, W	100 ± 26	26 ± 7	26
Leg press, N	744 ± 172	90 ± 30	12
Loaded stair-climb power, W	406 ± 77	56 ± 13	14

With data from Huang G, Basaria S, Travison TG, *et al.* Testosterone dose-response relationships in hysterectomized women with or without oophorectomy: effects on sexual function, body composition, muscle performance and physical function in a randomized trial. *Menopause* 2014;21:612–623. Data are shown as mean and SEM derived from Table 1 and digitized from Figure 4 from Huang *et al.* (112) showing the effects of testosterone (mean circulating concentration, 7.3 nmol/L) on muscle mass and strength in women treated with the highest testosterone dose (n = 11; 25 mg of testosterone enanthate per week).

(mean, 31 nmol/L) increased muscle mass by 19.2% (114). In a second study, 23 transmen administered adult male testosterone doses also produced striking increases in total body muscle size and limb muscle size (by 6.5% to 16.6%) and grip strength (by 18%) compared with age-matched untreated control women (115). Conversely, testosterone suppression (using an estrogen-based treatment regimen) in 20 transwomen (M2F transgender) that reduced circulating testosterone levels from adult male range to adult female range led to a 9.4% reduction in muscle mass (measured as cross-sectional area).

Effects on athletic performance

Muscle growth, as well as the increase in strength and power it brings, has an obvious performance-enhancing effect, in particular in sports that depend on strength and (explosive) power, such as track and field events (107, 110). There is convincing evidence that the sex differences in muscle mass and strength are sufficient to account for the increased strength and aerobic performance of men compared with women and is in keeping with the differences in world records between the sexes (116). The basis for the sex difference in muscle mass and strength is the sex difference in circulating testosterone as clearly shown (for example) by (1) the enhanced athletic performance of men compared with prepubertal boys and women (8); (2) the close correspondence of muscle growth (muscle size) with muscle strength in ascending dose studies in men by Bhasin *et al.* (111, 117–119) and Finkelstein *et al.* (65) and in postmenopausal women by Huang *et al.* (112); (3) the effect of male castration in reducing muscle size and strength, effects that are fully rectified by testosterone replacement; and (4) the striking efficacy of androgen doping on the sports performances of German Democratic Republic female athletes (120).

Hemoglobin

Biology

It is well known that levels of circulating hemoglobin are androgen-dependent and consequently higher in men than in women by 12% on average; however, the physiological mechanism by which androgens such as

testosterone boosts circulating hemoglobin is not fully understood (121). Testosterone increases secretion of and sensitivity to erythropoietin, the main trophic hormone for erythrocyte production and thereby hemoglobin synthesis, as well as suppressing hepcidin (122), a crucial iron regulatory protein that governs the body's iron economy. Hepcidin has to balance the need for iron absorption from foods (the only source of iron required for the body's iron-containing proteins) against the fact that the body has no mechanism to shed excess iron, which can be toxic. Adequate iron availability is essential for normal erythropoiesis and synthesis of key heme, iron-containing oxygen-transporting proteins such as hemoglobin and myoglobin (123) as well as other iron-dependent proteins such as cytochromes and DNA synthesis and repair enzymes. Experimental evidence in mice shows that testosterone increases myoglobin content of muscle with potential for augmenting aerobic exercise performance (96), but this has not been evaluated in humans.

Increasing the amount of hemoglobin in the blood has the biological effect of increasing oxygen transport from lungs to tissues, where the increased availability of oxygen enhances aerobic energy expenditure. This is exploited to its greatest effect in endurance sports (1). The experiments of Ekblom *et al.* (124) in 1972 (Fig. 4) demonstrated strong linear relationships between changes in hemoglobin [due to withdrawal or retransfusion of 1, 2 or 3 U (400 mL) of blood] and aerobic capacity, established by repeated testing of maximal exercise-induced oxygen consumption before and after each procedure (124). As already noted, circulating hemoglobin levels are on average 12% higher in men than women (125). It may be estimated that as a result the average maximal oxygen transfer will be ~10% greater in men than in women, which has a direct impact on their respective athletic capacities.

Observational data

The proposition that the sex difference in circulating hemoglobin levels is likely to be due to the sex difference in average circulating testosterone concentrations is supported by the fact that male castration (*e.g.*, for advanced prostate cancer) (126) and androgen deficiency due to reproductive system disorders (127) reduce circulating

hemoglobin in men, eliminating the sex difference, whereas testosterone replacement therapy restores circulating hemoglobin to adult male levels (121, 127, 128).

An unusually informative observational study of women with CAH provides unique insight into testosterone effects on circulating hemoglobin in otherwise healthy women (92). Women with CAH require glucocorticoid replacement therapy but exhibit widely varying levels of hormonal control (79). The degree of poor control is associated with increasing levels of circulating testosterone ranging from normal female concentrations up to 36 nmol/L, and these levels correlate closely ($r = 0.56$) with levels of circulating hemoglobin (Fig. 5). Interpolating from the dose-response regression, increases in circulating testosterone measured by LC-MS from 0.9 nmol/L to 5 nmol/L, 7 nmol/L, 10 nmol/L, and 19 nmol/L were associated with increases in circulating hemoglobin of 6.5%, 7.8%, 8.9%, and 11%, respectively, establishing a strong dose-response relationship. An 11% increase in circulating hemoglobin translates to a 10% difference in maximal oxygen transfer (124), which may account for virtually all the 12% sex difference in male and female circulating hemoglobin (125). To put this into context, any drug that achieved such increases in hemoglobin would be prohibited in sports for blood doping, as this difference is sufficient to have ergogenic effects, even without taking into account any testosterone effects on muscle mass or strength (for which data were not available in that study). Conversely, among elite female athletes with circulating testosterone in the healthy premenopausal female range, circulating hemoglobin does not correlate with athletic performance (35). In women with the mild hyperandrogenism of PCOS, circulating hemoglobin and hematocrit are reported as not (129) or marginally increased (130), findings that may be influenced by the fact that PCOS is

associated with reduced or absent menstruation, thereby reducing the iron loss of regular menstruation.

Interventional data

In the Bhasin *et al.* (111) studies, in both young and older men the highest testosterone dose produced a 12% increase in blood hemoglobin compared with the lowest dose, reflecting a strong dose-response relationship (Fig. 6) (131). Analogous findings were reported for testosterone treatment effects in postmenopausal women where the highest dose (25 mg weekly) of testosterone, which increased mean serum testosterone to 7.3 nmol/L, had the largest increase (3%) in blood hemoglobin and hematocrit (112).

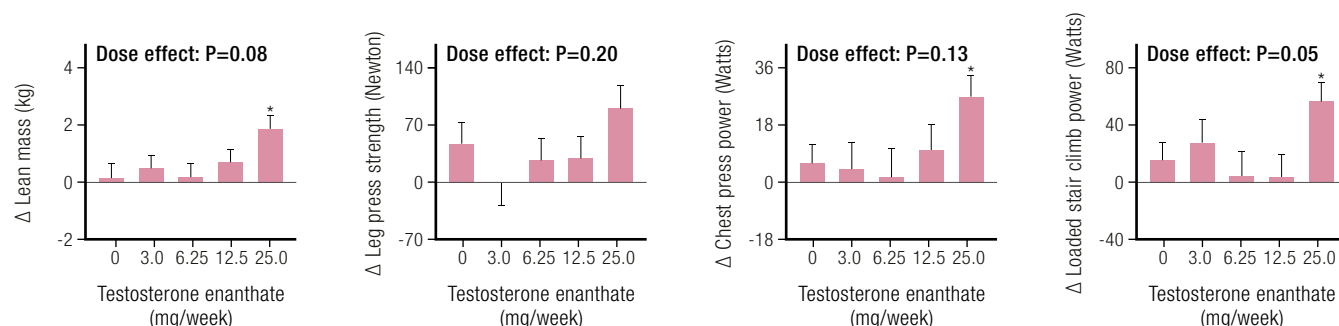
Corroborative findings are available from studies of transmen (F2M transgender), that is, natal females who subsequently receive testosterone treatment at replacement doses to create adult male circulating testosterone concentrations, who exhibit increases in circulating hemoglobin to male levels [reviewed in (132–134)]. Testosterone treatment in 17 (F2M) transmen that created mean circulating testosterone levels of 31 nmol/L also increased hemoglobin levels by 15% (114). Conversely, one prospective 12-month study of transgender (nonathlete) individuals reported that testosterone suppression (by an estrogen-based regimen) to normal female levels in 20 (M2F) transwomen reduced hemoglobin by 14%.

If such an increase in hemoglobin were produced by any chemical substance, it would be considered doping, according to the World Anti-Doping Code.

Bone

Biology

There is extensive experimental evidence from genetic mouse models showing that the sex differences in bone

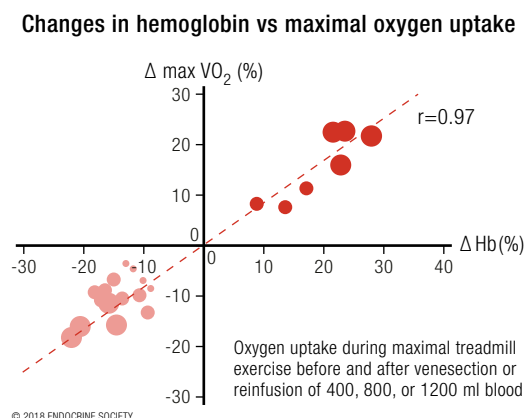


* Significant difference between mean on treatment change in dose group vs. placebo at 0.05 level. The significance level for the overall dose effect is by likelihood ratio test.

© 2018 ENDOCRINE SOCIETY

Figure 3. From Huang *et al.* (112): Dose-response effects on lean (muscle) mass and three measures of muscle strength as a result of increasing doses of weekly testosterone enanthate injections in women. Note the effects on all four parameters (three statistically significant) of the highest testosterone dose, the only one that produced circulating testosterone levels exceeding the normal female range. Reproduced with permission from Huang G, Basaria S, Travison TG, *et al.* Testosterone dose-response relationships in hysterectomized women with or without oophorectomy: effects on sexual function, body composition, muscle performance and physical function in a randomized trial. *Menopause* 2014;21:612–623.

Figure 4. Redrawn results from Ekblom *et al.* (124). Results from the transfusion of additional blood are shown in dark red circles and those after blood withdrawal in light red circles. Adapted with permission from Ekblom B, Goldbarg AN, Gullbring B. Response to exercise after blood loss and reinfusion. *J Appl Physiol* 1972;33:175–180.



size, mass, and function are due to the sex difference in circulating testosterone. These effects have been reported from studies of global and tissue or cell-selective inactivation of ARs or estrogen receptors that show that androgen effects are mediated by both direct effects on the AR as well as indirect effects mediated via aromatization of testosterone to estradiol to act on estrogen receptors [reviewed in (135)]. Bone grows in length due to epiphyseal chondral growth plates that provide cartilage, forming the matrix for lengthening of long bone, which is terminated by an estrogen-dependent mechanism that depends on aromatization of testosterone to estradiol. Similarly, bone width and density are increased through appositional growth from periosteal and endosteal expansion that depend on bone loading and androgen exposure together with other factors. An important difference between androgen effects on bone compared with effects on muscle or hemoglobin is that developmental bone effects of androgens are likely to be irreversible.

Observational data

Men have distinctively greater bone size, strength, and density than do women of the same age. As with muscle, sex differences in bone are absent prior to puberty but then accrue progressively from the onset of male puberty due to the sex difference in exposure to adult male circulating testosterone concentrations [reviewed in (135)]. The earlier onset of puberty and the related growth spurt in girls as well as earlier estrogen-dependent epiphyseal fusion explains shorter stature of girls than boys. As a result, on average men are 7% to 8% taller with longer, denser, and stronger bones, whereas women have shorter humerus and femur cross-sectional areas being 65% to 75% and 85%, respectively, those of men (106).

These changes create an advantage of greater bone strength and stronger fulcrum power from longer bones. Additionally, whereas passing through puberty enhances male physical performance, the widening of the female pelvis during puberty, balancing the evolutionary demands of obstetrics and locomotion (136, 137), retards the improvement in female physical performance, possibly driven by ovarian hormones rather than the absence of testosterone (138, 139).

Sex differences in height have been the most thoroughly investigated measure of bone size, as adult height is a stable, easily quantified measure in large population samples. Extensive twin studies show that adult height is highly heritable with predominantly additive genetic effects (140) that diverge in a sex-specific manner from the age of puberty onwards (141, 142), the effects of which are likely to be due to sex differences in adult circulating testosterone concentrations.

Bone density (total and medullary cross-sectional area) is increased in women with CAH with variably elevated serum testosterone (including into the male range) when it is only partially suppressed by glucocorticoid treatment (143), although more effective glucocorticoid suppression lowers bone density (144).

Interventional data

Well-designed, placebo-controlled direct interventional studies of supraphysiological androgen effects on bone in females are few, rarely feasible, and unlikely to be performed for ethical and practical reasons. Unlike muscle, which responds relatively rapidly to androgen effects so that muscle studies in humans can be completed within 3 to 4 months (65, 111, 112, 119, 145), comparable bone studies would typically take a year or more to reach plateau effects. Hence, such direct investigational studies in otherwise healthy women would risk side effects of virilization that may be only slowly and partly reversible, if at all, as well as potential promotion of hormone-dependent cancers making such studies ethically and practically not feasible.

Effects on athletic performance

The major effects of men's larger and stronger bones would be manifest via their taller stature as well as the larger fulcrum with greater leverage for muscular limb power exerted in jumping, throwing, or other explosive power activities. The greater cortical bone density and thereby resistance to long bone fractures is unlikely to be relevant to the athletic performance of young athletes, in whom fractures during competition are extremely rare and not expected to be linked to sex. Alternatively, stress fractures in athletes, mostly involving the legs, are more frequent in females with the male protection attributable to their larger and thicker bones (146).

Other androgen-sensitive sex dichotomous effects

Biology and observational data

Many if not most other aspects of physiology exhibit sex differences and may therefore enhance the impact of the male advantage in sports performance of the dominant determinants (muscle and hemoglobin). Examples include sex differences in exercise-induced cardiac (147, 148) and lung (149) function and mitochondrial biogenesis and energetics (95). However, the limited knowledge of the magnitude and hormonal mechanisms involved, specifically the degree of androgen dependence of these mechanisms, means that it is difficult to estimate their contribution, if any, toward the sex difference in athletic performance. The sex difference in pulmonary function may be largely explained by the androgen-sensitive sex difference in height, which is a strong predictor of lung capacity and function (149). Further physiological studies of the androgen dependence of other physiological sex differences are awaited with interest.

Psychological differences between men and women on mental function (e.g., rotational orientation) (150) as well as mood, motivation, and behavioral effects may involve androgen-sensitive effects during pre-natal and perinatal as well as postpubertal effects (151, 152).

Interventional data

There is some limited direct evidence from well-designed, placebo-controlled trials that administration of testosterone or other androgens at supraphysiological doses directly affect mood and behavior, notably inducing hypomania (153). In a randomized placebo-controlled study of testosterone administration in postmenopausal women (112), in case of those receiving the highest dose (the only one causing circulating testosterone levels to exceed the normal female range), there was not only an increase in muscle mass (4.4%) but a strikingly greater increase in muscle strength (12% to 26%), suggesting an enhanced mental motivational effect of testosterone on the effort-dependent tests of muscle strength.

Alternative Mechanisms Proposed to Explain Sex Differences in Athletic Performance

Alternative explanations for the sex difference in athletic performance, other than it being due to the sex difference in postpubertal circulating testosterone, have been proposed, including (1) sex differences in height because height is a predictor of muscle mass (116), (2) genetic sex differences due to the influence of unspecified Y chromosome genes (154), and (3) sex differences in GH secretion (116).

Effects of height

One proposal has been that, as men are taller than women, height differences may explain the sex differences in muscle mass and function, which explains some athletic success (116). Numerous factors contribute to the regulation of adult muscle mass, including genetics, race, adiposity, hormones, physical activity (exercise/training), diet, birth order, and bone size (including height) [reviewed in (155)]. Among the nonhormonal factors, genetics explains a large proportion [~50% to 60% from pooled twin studies (156)] of the variability in muscle mass and strength (157, 158) and may be explained in turn by the equally high genetic contributions to circulating testosterone (37, 38). Some factors influencing muscle mass and strength such as physical activity, adiposity, and bone size are also partly androgen-dependent. Prior to puberty there is no sex difference in skeletal features, including height (159, 160). However, with the onset of puberty, girls aged 11 and 12 years are transiently taller than peer-aged boys due to their earlier onset of the female pubertal growth spurt, but from the age of 14 years onward the taller stature in males emerges and stabilizes (141). Hence, similar to muscle mass, sex differences in bone size (including length, density, and height) arise after male puberty establishes the marked dichotomy between men and women in adult circulating testosterone concentrations. Taller height is

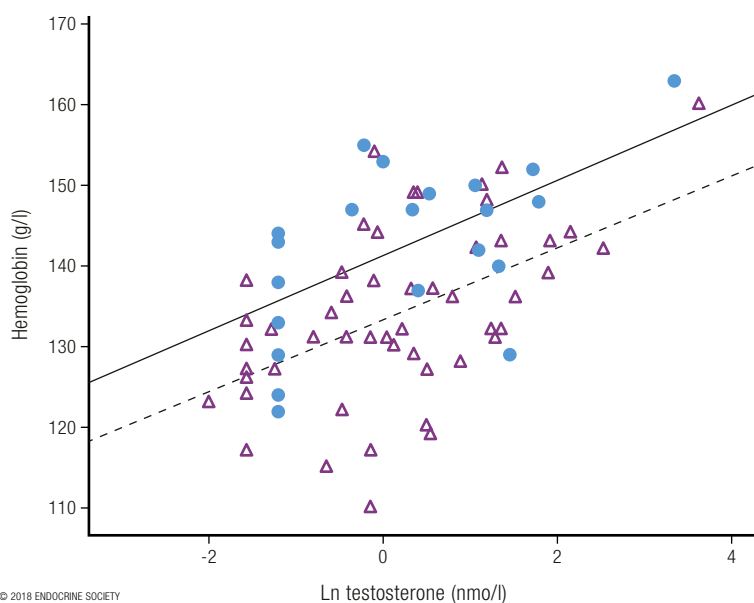


Figure 5. Plot of circulating hemoglobin against the natural logarithm of serum testosterone in women with congenital adrenal hyperplasia [from Karunasena *et al.* (92)]. The filled circles represent a cohort where serum testosterone was measured by immunoassay. The open triangles denote a second cohort, where serum testosterone was measured by LC-MS. Note the systematic overestimation of testosterone by the immunoassay used in cohort 1 vs LC-MS measurement in cohort 2. Despite that overestimation, however, the correlations were similar in both cohorts. Reproduced under a Creative Commons BY-NC-ND 4.0 license from Karunasena N, Han TS, Mallappa A, *et al.* Androgens correlate with increased erythropoiesis in women with congenital adrenal hyperplasia. Clin Endocrinol (Oxf) 2017;86:19–25.

advantageous in some sports (basketball, some football codes, combat sports), but in others (horse racing jockeys, cycling, gymnastics, weightlifting, bodybuilding) short stature provides a greater power/strength-to-weight ratio as well as superior rotational balance, speed, and agility. However, the male advantages in speed, strength, and endurance apply regardless of whether height is advantageous. Hence, the sex differences in height, where they exist, are largely dependent on postpubertal differences in circulating testosterone when sex differences in height are first expressed.

Genetic effects of Y chromosome

It has also been proposed that the sex difference in athletic performance may be due to genetic effects of an unspecified Y chromosome gene that may dictate taller stature (154), as height is correlated with men's greater muscle mass. The small human Y chromosome has few functional genes and none with a known effect on height other than the short stature homeobox (SHOX) gene, located in the pseudoautosomal regions of the tip of the short arms of X and Y chromosomes (161). Adult height displays an apparent dose dependency on SHOX gene copy number that is a major factor contributing to explaining both the short stature of 45,XO females (Turner syndrome), who have a single copy of the SHOX gene, as well as the tall stature of 47,XXY males (Klinefelter syndrome), who have three copies (161). However, when SHOX copy number is the same, men with additional supernumerary Y chromosomes (e.g., 47,XYY) are the same height as 47,XXY men (162). Hence, there is no evidence supporting dosage-dependent Y chromosomal gene effects on height independent of SHOX gene copy number, nor does men's possession of a Y chromosome explain the height difference between adult men and women. On the contrary, the tall stature of 47,XXY men is at least partly due to the concomitant androgen deficiency leading to pubertal

delay. Pubertal delay prolongs long bone growth due to delayed epiphyseal closure, an estrogen-dependent effect that requires adequate production of testosterone as a substrate for aromatization to estradiol, resulting in tall stature. Similar eunuchoidal features and taller stature are evident in 46,XY men with congenital hypogonadotropic hypogonadism (Kallmann syndrome and its variants) with comparable congenital onset of androgen deficiency, also manifest as pubertal delay and long bone overgrowth. Hence, taller height is better explained by impaired testicular function with delayed puberty and epiphyseal closure rather than unspecified Y chromosome dosage effects. In any case, rare aneuploidies in themselves do not explain the sex difference in height in the general population of individuals with normal sex chromosomes.

Growth hormone

The proposal that the sex difference in muscle mass and function might be due to sex differences in endogenous GH secretion (116) is refuted by the extensive and conclusive clinical evidence that endogenous GH secretion in young women is consistently higher (typically twice as high) as in young men of similar age (163–170). Those findings cannot explain the male advantage in muscle mass and strength unless GH retards muscle growth/function, for which there is no evidence. Furthermore, estrogens inhibit GH-dependent, hepatic IGF-1 production, the major pathway of GH action (171, 172). The weak observational association between low circulating IGF-1 and some, but not other, measures of weak muscle strength and limited mobility among older women may reflect general age-associated debility rather than any specific hormonal effects (173). Finally, the evidence that endogenous GH plays no role in sex differences in muscle mass and function is supported by evidence from the most extensive interventional study of GH treatment to non-GH-deficient adults, daily GH administration for 8 weeks to healthy recreational athletes produced only marginally significant improvement in exercise performance of men and none in women (174). These findings are consistent with the speculation that GH (or IGF-1) may be an amplifier of testosterone effects and therefore be a consequence of the sex difference in circulating testosterone rather than its cause.

The Impact of Adult Male Circulating Testosterone Concentrations on Sports Performance

Plausible estimates of the magnitude of the ergogenic advantage of adult male circulating testosterone concentrations are feasible from the limited available observational and interventional studies.

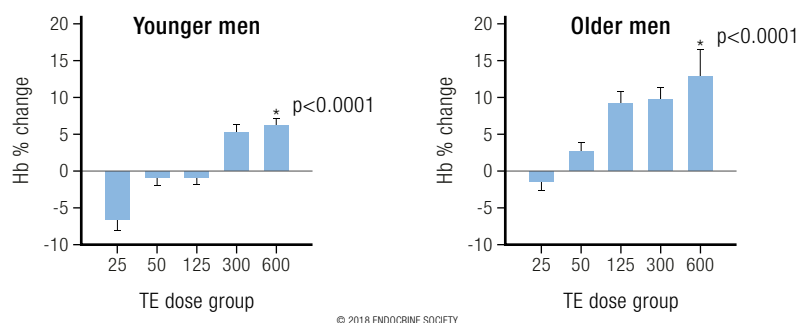


Figure 6. From Coviello *et al.* (131): Depicts the strong dose-response relationship between increasing testosterone dose with resulting change in blood hemoglobin in young and older men. Reproduced with permission from Coviello AD, Kaplan B, Lakshman KM, *et al.* Effects of graded doses of testosterone on erythropoiesis in healthy young and older men. *J Clin Endocrinol Metab* 2008;93:914–919.

Population data on the ontogeny of puberty show that prior to puberty boys and girls have comparable athletic performance, whereas sex differences in athletic performance emerge coinciding with the rise in circulating testosterone from the onset of male puberty. Male puberty results in circulating testosterone concentrations rising from the prepubertal and female postpubertal range (<2 nmol/L) to adult male circulating testosterone concentrations (18). This is associated with a 10% to 12% better performance in running and swimming events and 20% enhancement in jumping events (8).

A minimal estimate of the impact of adult male testosterone concentrations on muscle size and strength in females is provided by the Huang *et al.* (112) study of postmenopausal women. In this study the highest testosterone dose (weekly injections of 25 mg of testosterone enanthate) increased mean circulating testosterone from 0.9 nmol/L to 7.3 nmol/L, which is equivalent to the circulating testosterone of boys in early to middle puberty. After 24 weeks of testosterone treatment, the increase in circulating testosterone concentrations led to significant increases in muscle size of 4.4% and in muscle strength of 12% to 26%. Given the limited testosterone dose (and concentration) as well as study duration, it is likely that these findings underestimate the magnitude of the impact that sex difference in circulating testosterone has on muscle mass and strength, and therefore on athletic performance.

Converse effects of reduced athletic performance in athletes who undergo suppression of circulating testosterone concentrations from those in the male into the female range have been reported. Among recreational (nonelite) athletes, an observational study showed a consistent deterioration in athletic performance of transwomen (M2F transgender) athletes corresponding closely to the suppression of circulating testosterone concentrations (175). Similarly, among elite athletes with circulating testosterone in the male range due to DSDs, comparable findings of athletic performance reduced by an average of 5.7% when circulating testosterone was suppressed from the male range to <10 nmol/L (176). Subsequently, when the IAAF hyperandrogenism rule was suspended in 2015, and so these elite athletes could train and compete with unsuppressed serum testosterone levels, their athletic performances increased by a similar amount. Additionally, circulating hemoglobin levels in these untreated DSD athletes were comparable with male athletes or with female athletes doping with erythropoietin (Fig. 7). However, when circulating testosterone was suppressed to <10 nmol/L the levels of circulating hemoglobin were 12% lower and again comparable with nondoped, non-DSD females, corresponding to the 12% magnitude of the sex difference in hemoglobin between men and women (125).

Congruent findings are also known for an elite female athlete whose serial athletic performance based on publicly available best annual times between 2008 and 2016 for the 800-m running event are depicted in relationship to the original 2011 IAAF hyperandrogenism regulation (Fig. 8).

Based on the established dose-response relationships, suppression of circulating testosterone to <10 nmol/L would not eliminate all ergogenic benefits of testosterone for athletes competing in female events. For example, according to the Huang *et al.* (112) study, reducing circulating testosterone to a mean of 7.3 nmol/L would still deliver a 4.4% increase in muscle size and a 12% to 26% increase in muscle strength compared with circulating testosterone at the normal female mean value of 0.9 nmol/L. Similarly, according to the Karunasena *et al.* (92) study, reducing circulating testosterone concentration to 7 nmol/L would still deliver 7.8% more circulating hemoglobin than the normal female mean value. Hence, the magnitude of the athletic performance advantage in DSD athletes, which depends on the magnitude of elevated circulating testosterone concentrations, is considerably greater than the 5% to 9% difference observed in reducing levels to <10 nmol/L.

The physiological mechanism underlying these observations is further strengthened by prospective controlled studies of initiation of cross-sex hormone treatment in transgender individuals (114, 177). These show that during the first 12 months muscle mass (area) was decreased by 9.4% and hemoglobin levels by 14% in 20 transwomen (M2F transgender) treated with an estrogen-based regimen that reduced circulating testosterone concentrations from the male range to the female range. Conversely, in 17 transmen (F2M transgender) treated for the first time with testosterone for 12 months (which increased circulating testosterone levels to a mean of 31 nmol/L), muscle mass increased by 19.2% and hemoglobin by 15% (114). The muscle mass findings remained stable between 1 and 3 years after initiation of treatment, although fat mass continued to change between 1 and 3 years of testosterone treatment (177). These studies did not report muscle strength, but other studies of testosterone dose-response relationships for muscle mass and strength show consistently positive correlation (65, 93, 117, 119), although with disproportionately greater effect on muscle strength than on muscle mass. Hence, the muscle mass estimates in these prospective treatment initiation studies in transgender individuals likely underestimate the muscle strength gains from elevated testosterone levels where the circulating testosterone markedly exceeds female range to be within the male range as occurs in severe hyperandrogenism of DSD females, poorly controlled transwomen (M2F transgender), or transmen (F2M transgender). These effects are also the biological

basis of the ergogenic efficacy of androgen doping in women.

Finally, to put these competitive advantages into context, the winning margin (the difference in performance by which a competitor misses a gold medal, any medal, or making the final) in elite athletic or swimming events during the last three Olympics is <1% equally for both male and female events (Table 5).

Gaps in Knowledge and Research Limitations

The major limitations on scientific knowledge of the impact of adult male circulating testosterone concentrations on the sex difference in athletic performance is the lack of well-designed studies. Ideally, these would need to replicate adult male circulating testosterone concentrations for sufficient time in women to investigate the effects on muscle, hemoglobin, bone, and other androgen-sensitive measures that display consistent sex dichotomy in the population. However, the ethical and safety concerns preventing such studies hitherto are likely to remain formidable obstacles due to the risk of unacceptable and potentially irreversible virilization as well as of promoting hormone-dependent cancers in women.

With the exception of one interventional study administering a relatively low testosterone dose (*i.e.*, low for males) to women (112), the available evidence comprises observational studies that can only examine the effects of serum testosterone within physiological female limits or sparse and mostly uncontrolled data from intersex/DSD athletes. Although the available observational findings in healthy females are informative, the key question is the magnitude and dose

response of effects at still higher circulating testosterone concentrations on the performances of women. Whereas a testosterone dose-response relationship has been established in women at relatively low (for men) testosterone dose and circulating concentrations, it remains unproven (even if clearly plausible) that the testosterone dose-response relationships established in men for muscle, hemoglobin, and bone can be extrapolated to women when they are exposed to higher circulating testosterone concentrations (*i.e.*, comparable with male levels). It is theoretically possible there could be differences between men and women in muscle responses to testosterone, as muscle cell populations might express genetic differences in androgen sensitivity (for which there are no data), or alternatively the long-term prior pattern of testosterone exposure from conception to adulthood might lead to differences in testosterone dose responsiveness after maturity. Although the dose-response relationship in women may be similar to what is seen in men, there is also anecdotal evidence that the dose-response curves may be left shifted so that testosterone has greater potency in women than in men at comparable doses and circulating levels. The prediction is supported by the anecdotal evidence from the surreptitious East German national doping program in which the supervising doctors asserted from their experience of illicit cheating that androgens had more potent ergogenic effects in women than in men (120), a speculative opinion shared by many experienced sports medicine physicians.

There is no known means of increasing endogenous testosterone in women to anything like the requisite degree to attempt to answer these questions. In healthy men, circulating testosterone originates almost exclusively from a single source (testicular Leydig cells) and is subject to tight hypothalamic negative feedback control, so that either direct stimulation (by human chorionic gonadotropin) or indirect reflex effects (*e.g.*, from estrogen blockers operating via negative feedback) to enhance Leydig cell testosterone secretion are feasible. However, similar mechanisms do not operate in women, in whom circulating testosterone originates from three different sources (adrenal, ovary, extraglandular conversion of androgen precursors), none of which is subject to tight testosterone negative feedback control. As a result, it is not feasible to produce a sufficient increase in circulating testosterone in women either by direct ovarian stimulation or indirect reflex effects to test this hypothesis even if doing so were deemed ethical and safe. Alternatively, carefully controlled, graded-dose studies in F2M transgender individuals might be informative but are largely lacking at this time.

Hence, the only feasible design of such studies would be testosterone (or another androgen) administration to healthy young women. The only well-designed, placebo-controlled study of testosterone in

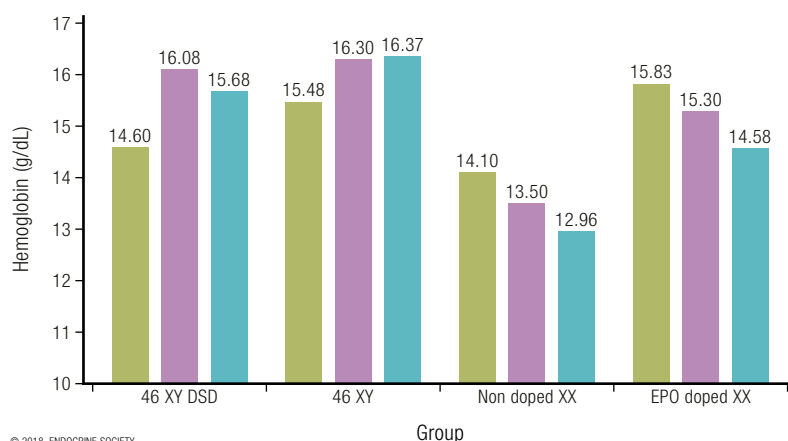


Figure 7. Mean hemoglobin concentrations (g/dL) of 12 elite athletes in 4 groups of 3 XY or XX middle-distance runners. The hemoglobin concentrations were collected as a part of the Athlete Biological Passport and analyzed according to the World Anti-Doping Agency standard methods. Each bar (athlete) is the mean of a minimum of three blood samples. In the 46,XY DSD group, blood was collected in a period when the athlete was not undergoing hormonal suppressive treatment.

otherwise healthy postmenopausal women was restricted to relatively low testosterone doses that, although clearly supraphysiological for women, were only 20% to 25% of male testosterone replacement doses (112). We are currently performing a double-blind, randomized, placebo-controlled study of the effects of moderately increased testosterone concentration on physical performance and behavior in young healthy women (ClinicalTrials.gov no. NCT03210558). However, obtaining ethical approval to administer supraphysiological testosterone doses that maintain circulating testosterone in the male range for sufficiently prolonged periods, as well as the practical difficulties in recruitment, are likely to remain obstacles to definitive resolution of this question.

In men, analogous ethical concerns over short- and long-term adverse effects delayed the definitive studies of supraphysiological testosterone doses to healthy young and older men but were eventually overcome. This was despite the fact that, uniquely among hormones, there is no known disease state in men due to pathologically excessive testosterone secretion. In contrast, in women, supraphysiological testosterone effects are known to produce virilization side effects that may be only slowly and partially, if at all, reversible. However, maintaining clearly supraphysiological testosterone concentrations would require treatment of months (muscle) or years (bone) and would replicate not only a known hyperandrogenic disease state (PCOS) but also potentially increasing risk of hormone-dependent cancers. In these circumstances, it could only be justifiable to replicate in women the salient testosterone dose-response studies available from men if the available evidence of dose-response relationship in men was not sufficiently convincing and/or there was reason to think that these dose-response characteristics would be substantially different in women. Overall, the unequivocal dose-response evidence in men together with the available overlap evidence in women appears sufficiently persuasive, so that it is doubtful that women would respond differently from men if their circulating testosterone levels were raised to the male range. More broadly, there is no more reason to require separate studies in women vs men than there is for every different ethnic subgroup of people. An aesthetic preference for splitting categories is not a sound reason to require the virtually impossible standard of establishing fresh and comprehensive empirical evidence in women of testosterone dose-response effects ranging into male circulating testosterone concentrations.

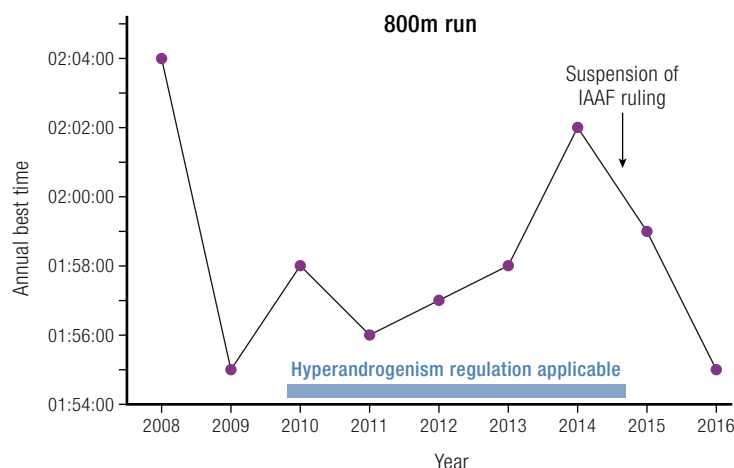
An analogy can be drawn to the World Anti-Doping Agency's practice of accepting salient surrogate evidence for banning the plethora of existing and new drugs with potential but individually unproven ergogenic effects where it is not feasible or ethical to require direct proof of the ergogenic effects. In that

context, the firmly established ergogenic efficacy of androgens (on muscle mass and strength) and increased hemoglobin (on endurance) [evidence reviewed in (1)] mean that chemical substances or methods that increase endogenous testosterone, erythropoietin, or hemoglobin are also considered ergogenic (178). By parity of reasoning, if a condition causes a female athlete's circulating testosterone levels to be in the male range, well exceeding normal female levels, with consequential increases in muscle, hemoglobin, and bone effects (at least), an ergogenic effect may reasonably be assumed.

Conclusions

The available, albeit incomplete, evidence makes it highly likely that the sex difference in circulating testosterone of adults explains most, if not all, the sex differences in sporting performance. This is based on the dose-response effects of circulating testosterone to increase muscle mass and strength, bone size and strength (density), and circulating hemoglobin, each of which alone increases athletic capacity, as well as other possible sex dichotomous, androgen-sensitive contributors such as mental effects (mood, motivation, aggression) and muscle myoglobin content. These facts explain the clear sex difference in athletic performance in most sports, on which basis it is commonly accepted that competition has to be divided into male and female categories.

The first IAAF hyperandrogenism regulation specified a hormonal eligibility criterion of a serum testosterone of <10 nmol/L for an androgen-sensitive athlete's participation in the protected category of female athletic events. This threshold was based on serum testosterone measurements by immunoassays.



© 2018 ENDOCRINE SOCIETY

Figure 8. Best annual 800-m times of an elite female athlete between 2008 and 2016. Data provided by Dr. Richard Auchus, University of Michigan, Ann Arbor, Michigan.

Table 5. The Winning Margin in Elite Athletic or Swimming Events During the Last Three Olympics

Median Margin (%) ^a	n	Win Gold	Win Medal	Make Final
Athletics ^b				
Running	81	0.62	0.31	0.22
Jumping	24	0.92	0.42	0.92
Throwing	24	1.93	0.70	0.75
Swimming ^c				
Backstroke	12	0.56	0.28	0.16
Breaststroke	12	0.84	0.14	0.17
Butterfly	12	0.52	0.48	0.12
Freestyle	30	0.49	0.23	0.14
Relay	18	0.37	0.35	0.12

^aWinning margin is defined as the difference (expressed as a percentage of the faster time) between first and second place (Win Gold), between third and fourth place (Win Medal), and between the last into the final and the first that missed out (Make Final). Years (2008, 2012, 2016) and sexes were combined as there were no significant differences in winning margin between them.

^bRunning includes 100 m, 200 m, 400 m, 800 m, 1500 m, 5000 m, 10,000 m, marathon, and 3000-m steeplechase, 110-m (male)/100-m (female) and 400-m hurdles, 4 × 100-m and 4 × 400-m relays, and 20-km and 50-km walk events. Jumping includes high jump, long jump, triple jump, and pole vault events. Throwing includes javelin, shot put, discus, and hammer events. Heptathlon and decathlon were not included as their final results are in points, not times.

^cEvents comprise 100 m and 200 m for the four strokes and 50 m, 100 m, 200 m, 400 m, 800 m (female)/1500 m (male) and marathon 10 km, with the relays being the 4 × 100-m medley and 4 × 100-m and 4 × 200-m freestyle relays.

However, no reliable method-independent consensus threshold could be established using commercial testosterone immunoassays, as these assays differ systematically due to method-specific bias arising unavoidably from the specificity of the different proprietary antibodies employed (25). Based on measurements using the more accurate and specific mass spectrometry methods, if the objective is to require female athletes with congenital conditions that cause them to have serum testosterone concentrations in the normal male range to bring those levels down to the same range as other female athletes, then (allowing for PCOS athletes) the threshold used should not be >5.0 nmol/L. This represents a conservative criterion that includes all healthy young (<40 years) women, including those with PCOS. Conversely, this criterion is generous to intersex/DSD females in allowing them to maintain a higher serum testosterone (2 to 5 nmol/L) than most non-PCOS competitors in female events even though increases in muscle mass and strength and hemoglobin would be expected in this range. This is so even though the range remains below the circulating testosterone levels of middle male puberty when the major biological effects of men's higher circulating testosterone begin to be fully expressed. Ongoing compliance with the eligibility criterion is also an important variable because the estrogen-based suppression of circulating testosterone, typically using daily administered estrogen products, has a rapid onset and offset. Adequate monitoring to prevent gaming of eligibility criteria would require

regular random rather than announced blood sampling.

A related matter is how long such a threshold of circulating testosterone should be maintained prior to competition. In both intersex/DSD and transgender individuals, the developmental effects of adult male circulating testosterone concentrations will have established the sex difference in muscle, hemoglobin, and bone, some of which is fixed and irreversible (bone size) and some of which is maintained by the male circulating testosterone concentrations (muscle, hemoglobin). The limited available prospective evidence from initiation of transgender cross-sex hormone treatment suggests that the advantageous increases in muscle and hemoglobin due to male circulating testosterone concentrations are induced or reversed during the first 12 months and the androgenic effects may plateau after time. This time course is much faster than the somatic effects of male puberty, which evolve over years and for some variables (e.g., peak bone mass) are not complete for up to a decade after the start of puberty. However, the abrupt hormonal changes induced by medical treatment in intersex/DSD or transgender individuals may be telescoped compared with male puberty where circulating testosterone concentrations increase irregularly and incompletely for some years. Additional data are available from the unique investigative model of men undergoing castration for prostate cancer. Just as androgen sensitivity to testosterone may differ between tissues (65), the time course of offset of

androgen effects following withdrawal of male testosterone concentrations may also differ between the major androgen-responsive tissues. For example, circulating hemoglobin shows a progressive fall for 6 months reaching a nadir and plateau at 12 to 16 months in six studies involving 534 men undergoing medical castration for prostate cancer (179–184). Although these studies of older men with prostate cancer must be extrapolated with caution, age, stage of disease, race, and baseline circulating

testosterone concentration did not affect the rate or extent of decline in hemoglobin (179, 181). Comparable longitudinal studies of muscle loss, strength, and performance following castration for prostate cancer are well summarized (185), showing progressive loss for 24 months (see Fig. 4). Further clinical studies to define the time course of changes, mainly offset, in testosterone-dependent effects, notably on muscle and hemoglobin, are badly needed to determine the optimal duration for cross-sex hormone effects in sports.

References

- Handelsman DJ. Performance enhancing hormones in sports doping. In: DeGroot LJ, Jameson JL, eds. *Endocrinology*. 7th ed. Philadelphia, PA: Elsevier Saunders; 2015:441–454.
- Coleman DL. Sex in sport. Available at: ssrn.com/abstract=2928106. Accessed 22 October 2017.
- Lee PA, Nordenström A, Houk CP, Ahmed SF, Auchus R, Baratz A, Baratz Dalke K, Liao LM, Lin-Su K, Looijenga LH III, Mazur T, Meyer-Bahlburg HF, Mouriquand P, Quigley CA, Sandberg DE, Vilain E, Witchel S; Global DSD Update Consortium. Global disorders of sex development update since 2006: perceptions, approach and care [published correction appears in *Horm Res Paediatr*. 2016;85(3):180]. *Horm Res Paediatr*. 2016;85(3):158–180.
- Southren AL, Tochimoto S, Carmody NC, Isurugi K. Plasma production rates of testosterone in normal adult men and women and in patients with the syndrome of feminizing testes. *J Clin Endocrinol Metab*. 1965;25(11):1441–1450.
- Horton R, Tait JF. Androstenedione production and interconversion rates measured in peripheral blood and studies on the possible site of its conversion to testosterone. *J Clin Invest*. 1966;45(3):301–313.
- Southren AL, Gordon GG, Tochimoto S. Further study of factors affecting the metabolic clearance rate of testosterone in man. *J Clin Endocrinol Metab*. 1968;28(8):1105–1112.
- Saez JM, Forest MG, Morera AM, Bertrand J. Metabolic clearance rate and blood production rate of testosterone and dihydrotestosterone in normal subjects, during pregnancy, and in hyperthyroidism. *J Clin Invest*. 1972;51(5):1226–1234.
- Handelsman DJ. Sex differences in athletic performance emerge coinciding with the onset of male puberty. *Clin Endocrinol (Oxf)*. 2017;87(1):68–72.
- Auchus RJ. Endocrinology and women's sports: the diagnosis matters. *Law Contemp Probl*. 2017;80:127–138.
- Foddy B, Savulescu J. Time to re-evaluate gender segregation in athletics? *Br J Sports Med*. 2011;45(15):1184–1188.
- Handelsman DJ. Androgen physiology, pharmacology and abuse. In: DeGroot LJ, Jameson JL, eds. *Endocrinology*. 7th ed. Philadelphia, PA: Elsevier Saunders; 2015:2368–2393.
- Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*. 2011;32(1):81–151.
- Abreu AP, Kaiser UB. Pubertal development and regulation. *Lancet Diabetes Endocrinol*. 2016;4(3):254–264.
- Horton R, Shinsako J, Forsham PH. Testosterone production and metabolic clearance rates with volumes of distribution in normal adult men and women. *Acta Endocrinol (Copenh)*. 1965;48:446–458.
- Rivarola MA, Saez JM, Meyer WJ, Jenkins ME, Migeon CJ. Metabolic clearance rate and blood production rate of testosterone and androst-4-ene-3,17-dione under basal conditions, ACTH and HCG stimulation. Comparison with urinary production rate of testosterone. *J Clin Endocrinol Metab*. 1966;26(11):1208–1218.
- Courant F, Aksglaede L, Antignac JP, Monteau F, Sorensen K, Andersson AM, Skakkebaek NE, Juul A, Bizec BL. Assessment of circulating sex steroid levels in prepubertal and pubertal boys and girls by a novel ultrasensitive gas chromatography-tandem mass spectrometry method. *J Clin Endocrinol Metab*. 2010;95(1):82–92.
- Davison SL, Bell R, Donath S, Montalto JC, Davis SR. Androgen levels in adult females: changes with age, menopause, and oophorectomy. *J Clin Endocrinol Metab*. 2005;90(7):3847–3853.
- Handelsman DJ, Sikaris K, Ly LP. Estimating age-specific trends in circulating testosterone and sex hormone-binding globulin in males and females across the lifespan. *Ann Clin Biochem*. 2016;53(Pt 3):377–384.
- Rothman MS, Carlson NE, Xu M, Wang C, Swerdloff R, Lee P, Goh VH, Ridgway EC, Wierman ME. Reexamination of testosterone, dihydrotestosterone, estradiol and estrone levels across the menstrual cycle and in postmenopausal women measured by liquid chromatography–tandem mass spectrometry. *Steroids*. 2011;76(1-2):177–182.
- Müller RK. History of doping and doping control. *Handb Exp Pharmacol*. 2010;(195):1–23.
- Rosner W, Hankinson SE, Sluss PM, Vesper HW, Wierman ME. Challenges to the measurement of estradiol: an Endocrine Society position statement. *J Clin Endocrinol Metab*. 2013;98(4):1376–1387.
- Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab*. 2007;92(2):405–413.
- Handelsman DJ, Wartofsky L. Requirement for mass spectrometry sex steroid assays in the *Journal of Clinical Endocrinology and Metabolism*. *J Clin Endocrinol Metab*. 2013;98(10):3971–3973.
- Handelsman DJ. Mass spectrometry, immunoassay and valid steroid measurements in reproductive medicine and science. *Hum Reprod*. 2017;32(6):1147–1150.
- Sikaris K, McLachlan RJ, Kazlauskas R, de Kretser D, Holden CA, Handelsman DJ. Reproductive hormone reference intervals for healthy fertile young men: evaluation of automated platform assays. *J Clin Endocrinol Metab*. 2005;90(11):5928–5936.
- Turpeinen U, Linko S, Itkonen O, Hämäläinen E. Determination of testosterone in serum by liquid chromatography–tandem mass spectrometry. *Scand J Clin Lab Invest*. 2008;68(1):50–57.
- Kushnir MM, Blamires T, Rockwood AL, Roberts WL, Yue B, Erdogan E, Bunker AM, Meikle AW. Liquid chromatography–tandem mass spectrometry assay for androstenedione, dehydroepiandrosterone, and testosterone with pediatric and adult reference intervals. *Clin Chem*. 2010;56(7):1138–1147.
- Salameh WA, Redor-Goldman MM, Clarke NJ, Reitz RE, Caulfield MP. Validation of a total testosterone assay using high-turbulence liquid chromatography tandem mass spectrometry: total and free testosterone reference ranges. *Steroids*. 2010;75(2):169–175.
- Neale SM, Hocking R, Biswas M, Turkes A, Rees D, Rees DA, Evans C. Adult testosterone and calculated free testosterone reference ranges by tandem mass spectrometry. *Ann Clin Biochem*. 2013;50(Pt 2):159–161.
- Kelsey TW, Li LQ, Mitchell RT, Whelan A, Anderson RA, Wallace WH. A validated age-related normative model for male total testosterone shows increasing variance but no decline after age 40 years [published correction appears in *PLoS One*. 2015;10(2):e0117674]. *PLoS One*. 2014;9(10):e109346.
- Hart RJ, Doherty DA, McLachlan RJ, Walls ML, Keelan JA, Dickinson JE, Skakkebaek NE, Norman RJ, Handelsman DJ. Testicular function in a birth cohort of young men. *Hum Reprod*. 2015;30(12):2713–2724.
- Travison TG, Vesper HW, Orwoll E, Wu F, Kaufman JM, Wang Y, Lapauw B, Fiers T, Matsumoto AM, Bhasin S. Harmonized reference ranges for circulating testosterone levels in men of four cohort studies in the United States and Europe. *J Clin Endocrinol Metab*. 2017;102(4):1161–1173.
- Haring R, Hannemann A, John U, Radke D, Nauck M, Wallaschofski H, Owen L, Adaway J, Keevil BG, Brabant G. Age-specific reference ranges for serum testosterone and androstenedione concentrations in women measured by liquid chromatography–tandem mass spectrometry. *J Clin Endocrinol Metab*. 2012;97(2):408–415.
- Bui HN, Sluss PM, Blynck S, Knol DL, Blankenstein MA, Heijboer AC. Dynamics of serum testosterone during the menstrual cycle evaluated by daily measurements with an ID-LC-MS/MS method and a 2nd generation automated immunoassay. *Steroids*. 2013;78(1):96–101.

35. Berman S, Garnier PY. Serum androgen levels and their relation to performance in track and field: mass spectrometry results from 2127 observations in male and female elite athletes. *Br J Sports Med.* 2017;**51**(17):1309–1314.
36. Eklund E, Berglund B, Labrie F, Carlström K, Ekström L, Hirschberg AL. Serum androgen profile and physical performance in women Olympic athletes. *Br J Sports Med.* 2017;**51**(17):1301–1308.
37. Travison TG, Zhuang WV, Lunetta KL, Karasik D, Bhasin S, Kiel DP, Coviello AD, Murabito JM. The heritability of circulating testosterone, oestradiol, oestrone and sex hormone binding globulin concentrations in men: the Framingham Heart Study. *Clin Endocrinol (Oxf).* 2014;**80**(2):277–282.
38. Coviello AD, Zhuang WV, Lunetta KL, Bhasin S, Ullor J, Zhang A, Karasik D, Kiel DP, Vasan RS, Murabito JM. Circulating testosterone and SHBG concentrations are heritable in women: the Framingham Heart Study. *J Clin Endocrinol Metab.* 2011;**96**(9):E1491–E1495.
39. Fui MN, Dupuis P, Grossmann M. Lowered testosterone in male obesity: mechanisms, morbidity and management. *Asian J Androl.* 2014;**16**(2):223–231.
40. Corona G, Rastrelli G, Monami M, Saad F, Luconi M, Luccchese M, Facchiano E, Sforza A, Forti G, Mannucci E, Maggi M. Body weight loss reverts obesity-associated hypogonadotropic hypogonadism: a systematic review and meta-analysis. *Eur J Endocrinol.* 2013;**168**(6):829–843.
41. Sartorius G, Spasevska S, Idan A, Turner L, Forbes E, Zamojska A, Allan CA, Ly LP, Conway AJ, McLachlan RI, Handelsman DJ. Serum testosterone, dihydrotestosterone and estradiol concentrations in older men self-reporting very good health: the healthy man study. *Clin Endocrinol (Oxf).* 2012;**77**(5):755–763.
42. Webb ML, Wallace JP, Hamill C, Hodgson JL, Mashaly MM. Serum testosterone concentration during two hours of moderate intensity treadmill running in trained men and women. *Endocr Res.* 1984;**10**(1):27–38.
43. Cano Sokoloff N, Misra M, Ackerman KE. Exercise, training, and the hypothalamic-pituitary-gonadal axis in men and women. *Front Horm Res.* 2016;**47**:27–43.
44. Bozdag G, Mumusoglu S, Zengin D, Karabulut E, Yildiz BO. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod.* 2016;**31**(12):2841–2855.
45. Hagmar M, Berglund B, Brismar K, Hirschberg AL. Hyperandrogenism may explain reproductive dysfunction in Olympic athletes. *Med Sci Sports Exerc.* 2009;**41**(6):1241–1248.
46. Eliakim A, Marom N, Galitskaya L, Nemet D. Hyperandrogenism among elite adolescent female athletes. *J Pediatr Endocrinol Metab.* 2010;**23**(8):755–758.
47. Rickenlund A, Carlström K, Ekblom B, Brismar TB, von Schoultz B, Hirschberg AL. Hyperandrogenicity is an alternative mechanism underlying oligomenorrhea or amenorrhea in female athletes and may improve physical performance. *Fertil Steril.* 2003;**79**(4):947–955.
48. Falhammar H, Nordenström A. Nonclassic congenital adrenal hyperplasia due to 21-hydroxylase deficiency: clinical presentation, diagnosis, treatment, and outcome. *Endocrine.* 2015;**50**(1):32–50.
49. Auchus RJ. The classic and nonclassic congenital adrenal hyperplasias. *Endocr Pract.* 2015;**21**(4):383–389.
50. Moran LJ, Munda PA, Teede HJ, Meikle PJ. The association of the lipidomic profile with features of polycystic ovary syndrome. *J Mol Endocrinol.* 2017;**59**(1):93–104.
51. Münzker J, Lindheim L, Adaway J, Trummer C, Lerchbaum E, Pieber TR, Keevil B, Obermayer-Pietsch B. High salivary testosterone-to-androstenedione ratio and adverse metabolic phenotypes in women with polycystic ovary syndrome. *Clin Endocrinol (Oxf).* 2017;**86**(4):567–575.
52. O'Reilly MW, Kempegowda P, Jenkinson C, Taylor AE, Quanson JL, Storbeck KH, Arlt W. 11-Oxygenated C19 steroids are the predominant androgens in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2017;**102**(3):840–848.
53. Handelsman DJ, Teede HJ, Desai R, Norman RJ, Moran LJ. Performance of mass spectrometry steroid profiling for diagnosis of polycystic ovary syndrome. *Hum Reprod.* 2017;**32**(2):418–422.
54. Pasquali R, Zanotti L, Fanelli F, Mezzullo M, Fazzini A, Morselli Labate AM, Repaci A, Ribichini D, Gambineri A. Defining hyperandrogenism in women with polycystic ovary syndrome: a challenging perspective. *J Clin Endocrinol Metab.* 2016;**101**(5):2013–2022.
55. Yang Y, Han Y, Wang W, Du T, Li Y, Zhang J, Yang D, Zhao X. Assessing new terminal body and facial hair growth during pregnancy: toward developing a simplified visual scoring system for hirsutism. *Fertil Steril.* 2016;**105**(2):494–500.
56. Tosi F, Fiers T, Kaufman JM, Dall'Alda M, Moretta R, Giagulli VA, Bonora E, Moghetti P. Implications of androgen assay accuracy in the phenotyping of women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 2016;**101**(2):610–618.
57. Daan NM, Jaspers L, Koster MP, Broekmans FJ, de Rijke YB, Franco OH, Laven JS, Kavousi M, Fauser BC. Androgen levels in women with various forms of ovarian dysfunction: associations with cardiometabolic features. *Hum Reprod.* 2015;**30**(10):2376–2386.
58. Bui HN, Sluss PM, Hayes FJ, Blincko S, Knol DL, Blankenstein MA, Heijboer AC. Testosterone, free testosterone, and free androgen index in women: reference intervals, biological variation, and diagnostic value in polycystic ovary syndrome. *Clin Chim Acta.* 2015;**450**:227–232.
59. Keefe CC, Goldman MM, Zhang K, Clarke N, Reitz RE, Welt CK. Simultaneous measurement of thirteen steroid hormones in women with polycystic ovary syndrome and control women using liquid chromatography-tandem mass spectrometry. *PLoS One.* 2014;**9**(4):e93805.
60. Yasmin E, Balen AH, Barth JH. The association of body mass index and biochemical hyperandrogenaemia in women with and without polycystic ovary syndrome. *Eur J Obstet Gynecol Reprod Biol.* 2013;**166**(2):173–177.
61. Janse F, Eijkemans MJ, Goverde AJ, Lentjes EG, Hoek A, Lambalk CB, Hickey TE, Fauser BC, Norman RJ. Assessment of androgen concentration in women: liquid chromatography-tandem mass spectrometry and extraction RIA show comparable results. *Eur J Endocrinol.* 2011;**165**(6):925–933.
62. Jedel E, Gustafson D, Waern M, Sverrisdottir YB, Landén M, Janson PO, Labrie F, Ohlsson C, Stener-Victorin E. Sex steroids, insulin sensitivity and sympathetic nerve activity in relation to affective symptoms in women with polycystic ovary syndrome. *Psychoneuroendocrinology.* 2011;**36**(10):1470–1479.
63. Legro RS, Schlaff WD, Diamond MP, Coutifaris C, Casson PR, Brzyski RG, Christman GM, Trussell JC, Krawetz SA, Snyder PJ, Ohl D, Carson SA, Steinkampf MP, Carr BR, McGovern PG, Cataldo NA, Gosman CG, Nestler JE, Myers ER, Santoro N, Eisenberg E, Zhang M, Zhang H; Reproductive Medicine Network. Total testosterone assays in women with polycystic ovary syndrome: precision and correlation with hirsutism. *J Clin Endocrinol Metab.* 2010;**95**(12):5305–5313.
64. Stener-Victorin E, Holm G, Labrie F, Nilsson L, Janson PO, Ohlsson C. Are there any sensitive and specific sex steroid markers for polycystic ovary syndrome? *J Clin Endocrinol Metab.* 2010;**95**(2):810–819.
65. Finkelstein JS, Lee H, Burnett-Bowie SA, Pallais JC, Yu EW, Borges LF, Jones BF, Barry CV, Wulczyn KE, Thomas BJ, Leder BZ. Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med.* 2013;**369**(11):1011–1022.
66. Donovan KA, Gonzalez BD, Nelson AM, Fishman MN, Zachariah B, Jacobsen PB. Effect of androgen deprivation therapy on sexual function and bother in men with prostate cancer: a controlled comparison. *Psychooncology.* 2018;**27**(1):316–324.
67. Buena F, Swerdloff RS, Steiner BS, Lutchmansingh P, Peterson MA, Pandian MR, Galmirani M, Bhasin S. Sexual function does not change when serum testosterone levels are pharmacologically varied within the normal male range. *Fertil Steril.* 1993;**59**(5):1118–1123.
68. Sartorius GA, Ly LP, Handelsman DJ. Male sexual function can be maintained without aromatization: randomized placebo-controlled trial of dihydrotestosterone (DHT) in healthy, older men for 24 months. *J Sex Med.* 2014;**11**(10):2562–2570.
69. Liu PY, Swerdloff RS, Christenson PD, Handelsman DJ, Wang C; Hormonal Male Contraception Summit Group. Rate, extent, and modifiers of spermatogenic recovery after hormonal male contraception: an integrated analysis. *Lancet.* 2006;**367**(9520):1412–1420.
70. Walsh PC, Swerdloff RS. Biphasic effect of testosterone on spermatogenesis in the rat. *Invest Urol.* 1973;**11**(3):190–193.
71. Singh J, O'Neill C, Handelsman DJ. Induction of spermatogenesis by androgens in gonadotropin-deficient (hpg) mice. *Endocrinology.* 1995;**136**(12):5311–5321.
72. Handelsman DJ, Spaliviero JA, Simpson JM, Allan CM, Singh J. Spermatogenesis without gonadotropins: maintenance has a lower testosterone threshold than initiation. *Endocrinology.* 1999;**140**(9):3938–3946.
73. Juel Mortensen L, Blomberg Jensen M, Christiansen P, Rønholt AM, Jørgensen A, Frederiksen H, Nielsen JE, Loya AC, Grønkær Toft B, Skakkebaek NE, Rajpert-De Meyts E, Juul A. Germ cell neoplasia in situ and preserved fertility despite suppressed gonadotropins in a patient with testotoxicosis. *J Clin Endocrinol Metab.* 2017;**102**(12):4411–4416.
74. Cunha-Silva M, Brito VN, Macedo DB, Bessa DS, Ramos CO, Lima LG, Barroso PS, Arnhold IJP, Segaloff DL, Mendonça BB, Latronico AC. Spontaneous fertility in a male patient with testotoxicosis despite suppression of FSH levels. *Hum Reprod.* 2018;**33**(5):914–918.
75. Mendonça BB, Batista RL, Domenice S, Costa EM, Arnhold IJ, Russell DW, Wilson JD. Steroid 5 α -reductase 2 deficiency. *J Steroid Biochem Mol Biol.* 2016;**163**:206–211.
76. Mendonça BB, Gomes NL, Costa EM, Inacio M, Martin RM, Nishi MY, Carvalho FM, Tibor FD, Domenice S. 46,XY disorder of sex development (DSD) due to 17 β -hydroxysteroid dehydrogenase type 3 deficiency. *J Steroid Biochem Mol Biol.* 2017;**165**(Pt A):79–85.

77. Quigley CA, De Bellis A, Marschke KB, el-Awady MK, Wilson EM, French FS. Androgen receptor defects: historical, clinical, and molecular perspectives. *Endocr Rev.* 1995;**16**(3):271–321.
78. Lucas-Herald A, Bertelloni S, Juul A, Bryce J, Jiang J, Rodie M, Sinnott R, Boroujerdi M, Lindhardt Johansen M, Hiort O, Holterhus PM, Cools M, Guaragna-Filho G, Guerra-Junior G, Weintrob N, Hannema S, Drop S, Guran T, Darendeliler F, Nordenstrom A, Hughes IA, Acerini C, Tadokoro-Cuccaro R, Ahmed SF. The long-term outcome of boys with partial androgen insensitivity syndrome and a mutation in the androgen receptor gene. *J Clin Endocrinol Metab.* 2016;**101**(11):3959–3967.
79. El-Maouche D, Artl W, Merke DP. Congenital adrenal hyperplasia. *Lancet.* 2017;**390**(10108):2194–2210.
80. Bermon S, Garnier PY, Hirschberg AL, Robinson N, Giraud S, Nicoli R, Baume N, Saugy M, Fénichel P, Bruce SJ, Henry H, Dollé G, Ritzen M. Serum androgen levels in elite female athletes. *J Clin Endocrinol Metab.* 2014;**99**(11):4328–4335.
81. Imperato-McGinley J, Peterson RE, Gautier T, Sturla E. Androgens and the evolution of male-gender identity among male pseudohermaphrodites with 5 α -reductase deficiency. *N Engl J Med.* 1979;**300**(22):1233–1237.
82. Kang HJ, Imperato-McGinley J, Zhu YS, Rosenwaks Z. The effect of 5 α -reductase-2 deficiency on human fertility. *Fertil Steril.* 2014;**101**(2):310–316.
83. Strickland AL, French FS. Absence of response to dihydrotestosterone in the syndrome of testicular feminization. *J Clin Endocrinol Metab.* 1969;**29**(9):1284–1286.
84. Rosenfield RL, Lawrence AM, Liao S, Landau RL. Androgens and androgen responsiveness in the feminizing testis syndrome. Comparison of complete and “incomplete” forms. *J Clin Endocrinol Metab.* 1971;**32**(5):625–632.
85. Hamilton CR Jr, Kliman B. Anabolic effect of dihydrotestosterone in testicular feminization syndrome. *Metabolism.* 1971;**20**(9):870–877.
86. Zachmann M, Zagalak M, Völlmin J, Gitzelmann RP, Prader A. Influence of testosterone on urinary ¹⁵N-balance in normal subjects and patients with testicular feminization. *Clin Chim Acta.* 1977;**77**(2):147–157.
87. Tincello DG, Saunders PT, Hodgins MB, Simpson NB, Edwards CR, Hargreaves TB, Wu FC. Correlation of clinical, endocrine and molecular abnormalities with in vivo responses to high-dose testosterone in patients with partial androgen insensitivity syndrome. *Clin Endocrinol (Oxf).* 1997;**46**(4):497–506.
88. Grino PB, Isidro-Gutierrez RF, Griffin JE, Wilson JD. Androgen resistance associated with a qualitative abnormality of the androgen receptor and responsive to high dose androgen therapy. *J Clin Endocrinol Metab.* 1989;**68**(3):578–584.
89. Lundberg Giwerzman Y, Nikoshkov A, Lindsten K, Byström B, Pousette A, Knudtzon J, Alm J, Wedell A. Response to treatment in patients with partial androgen insensitivity due to mutations in the DNA-binding domain of the androgen receptor. *Horm Res.* 2000;**53**(2):83–88.
90. Holterhus PM, Sinnecker GH, Hiort O. Phenotypic diversity and testosterone-induced normalization of mutant L712F androgen receptor function in a kindred with androgen insensitivity. *J Clin Endocrinol Metab.* 2000;**85**(9):3245–3250.
91. Quigley CA. *The androgen receptor: physiology and pathophysiology.* In: Nieschlag E, Behre HM, eds. *Testosterone: Action, Deficiency, Substitution.* 2nd ed. Berlin, Germany: Springer-Verlag; 1998:33–106.
92. Karunasena N, Han TS, Mallappa A, Elman M, Merke DP, Ross RJ, Daniel E. Androgens correlate with increased erythropoiesis in women with congenital adrenal hyperplasia. *Clin Endocrinol (Oxf).* 2017;**86**(1):19–25.
93. Herbst KL, Bhasin S. Testosterone action on skeletal muscle. *Curr Opin Clin Nutr Metab Care.* 2004;**7**(3):271–277.
94. Dubois V, Laurent MR, Sinnesael M, Cielien N, Helsen C, Clinckemalie L, Spans L, Gayan-Ramirez G, Deldicque L, Hespel P, Carmeliet G, Vanderschueren D, Claessens F. A satellite cell-specific knockout of the androgen receptor reveals myostatin as a direct androgen target in skeletal muscle. *FASEB J.* 2014;**28**(7):2979–2994.
95. Usui T, Kajita K, Kajita T, Mori I, Hanamoto T, Ikeda T, Okada H, Taguchi K, Kitada Y, Morita H, Sasaki T, Kitamura T, Sato T, Kojima I, Ishizuka T. Elevated mitochondrial biogenesis in skeletal muscle is associated with testosterone-induced body weight loss in male mice. *FEBS Lett.* 2014;**588**(10):1935–1941.
96. Mänttari S, Anttila K, Järviö M. Testosterone stimulates myoglobin expression in different muscles of the mouse. *J Comp Physiol B.* 2008;**178**(7):899–907.
97. Ferrando AA, Sheffield-Moore M, Yeckel CW, Gilkison C, Jiang J, Achacosa A, Lieberman SA, Tipton K, Wolfe RR, Urban RJ. Testosterone administration to older men improves muscle function: molecular and physiological mechanisms. *Am J Physiol Endocrinol Metab.* 2002;**282**(3):E601–E607.
98. Matzuk MM, Lamb DJ. The biology of infertility: research advances and clinical challenges. *Nat Med.* 2008;**14**(11):1197–1213.
99. Matzuk MM, Lamb DJ. Genetic dissection of mammalian fertility pathways. *Nat Cell Biol.* 2002;**4**(Suppl):S41–S49.
100. Walters KA, Simanainen U, Handelsman DJ. Molecular insights into androgen actions in male and female reproductive function from androgen receptor knockout models. *Hum Reprod Update.* 2010;**16**(5):543–558.
101. MacLean HE, Chiu WS, Notini AJ, Axell AM, Davey RA, McManus JF, Ma C, Plant DR, Lynch GS, Zajac JD. Impaired skeletal muscle development and function in male, but not female, genomic androgen receptor knockout mice. *FASEB J.* 2008;**22**(8):2676–2689.
102. Morrow JR Jr, Hosler WW. Strength comparisons in untrained men and trained women athletes. *Med Sci Sports Exerc.* 1981;**13**(3):194–197.
103. Miller AE, MacDougall JD, Tarnopolsky MA, Sale DG. Gender differences in strength and muscle fiber characteristics. *Eur J Appl Physiol Occup Physiol.* 1993;**66**(3):254–262.
104. Janssen I, Heymsfield SB, Wang ZM, Ross R. Skeletal muscle mass and distribution in 468 men and women aged 18–88 yr. *J Appl Physiol.* 2000;**89**(1):81–88.
105. Hosler WW, Morrow JR Jr. Arm and leg strength compared between young women and men after allowing for differences in body size and composition. *Ergonomics.* 1982;**25**(4):309–313.
106. Sale DG. Neuromuscular function. In: Tarnopolsky M, ed. *Gender Differences in Metabolism: Practical and Nutritional Implications.* Boca Raton, FL: CRC Press; 1999:61–86.
107. Tønnessen E, Svendsen IS, Olsen IC, Guttormsen A, Haugen T. Performance development in adolescent track and field athletes according to age, sex and sport discipline. *PLoS One.* 2015;**10**(6):e0129014.
108. Carmina E, Guastella E, Longo RA, Rini GB, Lobo RA. Correlates of increased lean muscle mass in women with polycystic ovary syndrome. *Eur J Endocrinol.* 2009;**161**(4):583–589.
109. Douchi T, Oki T, Yamasaki H, Kuwahata R, Nakae M, Nagata Y. Relationship of androgens to muscle size and bone mineral density in women with polycystic ovary syndrome. *Obstet Gynecol.* 2001;**98**(3):445–449.
110. Cardinale M, Stone MH. Is testosterone influencing explosive performance? *J Strength Cond Res.* 2006;**20**(1):103–107.
111. Bhasin S, Woodhouse L, Casaburi R, Singh AB, Bhasin D, Berman N, Chen X, Yarasheski KE, Magliano L, Dzekov C, Dzekov J, Bross R, Phillips J, Sinha-Hikim I, Shen R, Storer TW. Testosterone dose-response relationships in healthy young men. *Am J Physiol Endocrinol Metab.* 2001;**281**(6):E1172–E1181.
112. Huang G, Basaria S, Travison TG, Ho MH, Davda M, Mazer NA, Miciek R, Knapp PE, Zhang A, Collins L, Ursino M, Appleman E, Dzekov C, Stroh H, Ouellette M, Rundell T, Baby M, Bhatia NN, Khorram O, Friedman T, Storer TW, Bhasin S. Testosterone dose-response relationships in hysterectomized women with or without oophorectomy: effects on sexual function, body composition, muscle performance and physical function in a randomized trial. *Menopause.* 2014;**21**(6):612–623.
113. Dobs AS, Nguyen T, Pace C, Roberts CP. Differential effects of oral estrogen versus oral estrogen-androgen replacement therapy on body composition in postmenopausal women. *J Clin Endocrinol Metab.* 2002;**87**(4):1509–1516.
114. Elbers JM, Asscheman H, Seidell JC, Gooren LJ. Effects of sex steroid hormones on regional fat depots as assessed by magnetic resonance imaging in transsexuals. *Am J Physiol.* 1999;**276**(2 Pt 1):E317–E325.
115. Van Caenegem E, Wierckx K, Taes Y, Schreiner T, Vandewalle S, Toye K, Lapauw B, Kaufman JM, T'Sjoen G. Body composition, bone turnover, and bone mass in trans men during testosterone treatment: 1-year follow-up data from a prospective case-controlled study (ENIGI). *Eur J Endocrinol.* 2015;**172**(2):163–171.
116. Sonksen P. Determination and regulation of body composition in elite athletes. *Br J Sports Med.* 2018;**52**(4):219–229.
117. Storer TW, Woodhouse L, Magliano L, Singh AB, Dzekov C, Dzekov J, Bhasin S. Changes in muscle mass, muscle strength, and power but not physical function are related to testosterone dose in healthy older men. *J Am Geriatr Soc.* 2008;**56**(11):1991–1999.
118. Bhasin S, Parker RA, Sattler F, Haubrich R, Alston B, Umbleja T, Shikuma CM; AIDS Clinical Trials Group Protocol A5079 Study Team. Effects of testosterone supplementation on whole body and regional fat mass and distribution in human immunodeficiency virus-infected men with abdominal obesity. *J Clin Endocrinol Metab.* 2007;**92**(3):1049–1057.
119. Bhasin S, Woodhouse L, Casaburi R, Singh AB, Mac RP, Lee M, Yarasheski KE, Sinha-Hikim I, Dzekov C, Dzekov J, Magliano L, Storer TW. Older men are as responsive as young men to the anabolic effects of graded doses of testosterone on the skeletal muscle. *J Clin Endocrinol Metab.* 2005;**90**(2):678–688.
120. Franke WW, Berendonk B. Hormonal doping and androgenization of athletes: a secret program of the German Democratic Republic government. *Clin Chem.* 1997;**43**(7):1262–1279.
121. Shahani S, Braga-Basaria M, Maggio M, Basaria S. Androgens and erythropoiesis: past and present. *J Endocrinol Invest.* 2009;**32**(8):704–716.
122. Bachman E, Travison TG, Basaria S, Davda MN, Guo W, Li M, Connor Westfall J, Bae H, Gondeuk V,

- Bhasin S. Testosterone induces erythrocytosis via increased erythropoietin and suppressed hepcidin: evidence for a new erythropoietin/hemoglobin set point. *J Gerontol A Biol Sci Med Sci*. 2014;**69**(6): 725–735.
123. Ordway GA, Garry DJ. Myoglobin: an essential hemoprotein in striated muscle. *J Exp Biol*. 2004; **207**(Pt 20):3441–3446.
 124. Ekblom B, Goldbarg AN, Gullbring B. Response to exercise after blood loss and reinfusion. *J Appl Physiol*. 1972;**33**(2):175–180.
 125. Murphy WG. The sex difference in haemoglobin levels in adults—mechanisms, causes, and consequences. *Blood Rev*. 2014;**28**(2):41–47.
 126. Grossmann M, Zajac JD. Hematological changes during androgen deprivation therapy. *Asian J Androl*. 2012;**14**(2):187–192.
 127. Snyder PJ, Peachey H, Berlin JA, Hannoush P, Haddad G, Dlewati A, Santanna J, Loh L, Lenrow DA, Holmes JH, Kapoor SC, Atkinson LE, Strom BL. Effects of testosterone replacement in hypogonadal men. *J Clin Endocrinol Metab*. 2000;**85**(8):2670–2677.
 128. Roy CN, Snyder PJ, Stephens-Shields AJ, Artz AS, Bhasin S, Cohen HJ, Farrar JT, Gill TM, Zeldow B, Cella D, Barrett-Connor E, Cauley JA, Crandall JP, Cunningham GR, Ensrud KE, Lewis CE, Matsumoto AM, Molitch ME, Pahor M, Swerdloff RS, Cifelli D, Hou X, Resnick SM, Walston JD, Anton S, Basaria S, Diem SJ, Wang C, Schrier SL, Ellenberg SS. Association of testosterone levels with anemia in older men: a controlled clinical trial. *JAMA Intern Med*. 2017;**177**(4):480–490.
 129. Berria R, Gastaldelli A, Lucidi S, Belfort R, De Filippis E, Easton C, Brytzki R, Cusi K, Jovanovic L, DeFronzo R. Reduction in hematoctrit level after pioglitazone treatment is correlated with decreased plasma free testosterone level, not hemodilution, in women with polycystic ovary syndrome. *Clin Pharmacol Ther*. 2006;**80**(2):105–114.
 130. Han Y, Kim HS, Lee HJ, Oh JY, Sung YA. Metabolic effects of polycystic ovary syndrome in adolescents. *Ann Pediatr Endocrinol Metab*. 2015;**20**(3): 136–142.
 131. Coviello AD, Kaplan B, Lakshman KM, Chen T, Singh AB, Bhasin S. Effects of graded doses of testosterone on erythropoiesis in healthy young and older men. *J Clin Endocrinol Metab*. 2008;**93**(3):914–919.
 132. Irwig MS. Testosterone therapy for transgender men. *Lancet Diabetes Endocrinol*. 2017;**5**(4):301–311.
 133. Velho I, Figuera TM, Ziegelmann PK, Spritzer PM. Effects of testosterone therapy on BMI, blood pressure, and laboratory profile of transgender men: a systematic review. *Andrology*. 2017;**5**(5): 881–888.
 134. Jacobsen JW, Gooren LJ, Schulte HM. Safety aspects of 36 months of administration of long-acting intramuscular testosterone undecanoate for treatment of female-to-male transgender individuals. *Eur J Endocrinol*. 2009;**161**(5):795–798.
 135. Almeida M, Laurent MR, Dubois V, Claessens F, O'Brien CA, Bouillon R, Vanderschueren D, Manolagas SC. Estrogens and androgens in skeletal physiology and pathophysiology. *Physiol Rev*. 2017; **97**(1):135–187.
 136. Sharma K, Gupta P, Shandilya S. Age related changes in pelvis size among adolescent and adult females with reference to parturition from Naraingarh, Haryana (India). *Homo*. 2016;**67**(4):273–293.
 137. Fischer B, Mitteroecker P. Allometry and sexual dimorphism in the human pelvis. *Anat Rec (Hoboken)*. 2017;**300**(4):698–705.
 138. Riesenfeld A. Functional and hormonal control of pelvic width in the rat. *Acta Anat (Basel)*. 1978; **102**(4):427–432.
 139. Berdnikovs S, Bernstein M, Metzler A, German RZ. Pelvic growth: ontogeny of size and shape sexual dimorphism in rat pelvis. *J Morphol*. 2007;**268**(1): 12–22.
 140. Polderman TJ, Benjamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*. 2015; **47**(7):702–709.
 141. Jelenkovic A, Sund R, Hur YM, Yokoyama Y, Hjelmberg JV, Möller S, Honda C, Magnusson PK, Pedersen NL, Ooki S, Aaltonen S, Stazi MA, Fagnani C, D'Ippolito C, Freitas DL, Maia JA, Ji F, Ning F, Pang Z, Rebato E, Busjahn A, Kandler C, Saudino KJ, Jang KL, Cozen W, Hwang AE, Mack TM, Gao W, Yu C, Li L, Corley RP, Huibregtse BM, Derom CA, Vlietinck RF, Loos RJ, Heikkilä K, Wardle J, Llewellyn CH, Fisher A, McAdams TA, Eley TC, Gregory AM, He M, Ding X, Bjerregaard-Andersen M, Beck-Nielsen H, Sodemann M, Tarnoki AD, Tarnoki DL, Knafo-Noam A, Mankuta D, Abramson L, Burt SA, Klump KL, Silberg JL, Eaves LJ, Maes HH, Krueger RF, McGue M, Pahlen S, Gatz M, Butler DA, Bartels M, van Beijsterveldt TC, Craig JM, Saffery R, Dubois L, Boivin M, Brendgen M, Dionne G, Vitaro F, Martin NG, Medland SE, Montgomery GW, Swan GE, Krasnow R, Tynelius P, Lichtenstein P, Haworth CM, Plomin R, Bayasgalan G, Narandalai D, Harden KP, Tucker-Drob EM, Spector T, Mangino M, Lachance G, Baker LA, Tuvblad C, Duncan GE, Buchwald D, Willemsen G, Skytthe A, Kyvik KO, Christensen K, Öncel SY, Aliev F, Rasmussen F, Goldberg JH, Sørensen TI, Boomsma DI, Kaprio J, Silventoinen K. Genetic and environmental influences on height from infancy to early adulthood: an individual-based pooled analysis of 45 twin cohorts. *Sci Rep*. 2016;**6**(1):28496.
 142. Jelenkovic A, Hur YM, Sund R, Yokoyama Y, Siribaddana SH, Hotopf M, Sumathipala A, Rijdsdijk F, Tan Q, Zhang D, Pang Z, Aaltonen S, Heikkilä K, Öncel SY, Aliev F, Rebato E, Tarnoki AD, Tarnoki DL, Christensen K, Skytthe A, Kyvik KO, Silberg JL, Eaves LJ, Maes HH, Cutler TL, Hopper JL, Ordoñana JR, Sánchez-Romera JF, Colodro-Conde L, Cozen W, Hwang AE, Mack TM, Sung J, Song YM, Yang S, Lee K, Franz CE, Kremen WS, Lyons MJ, Busjahn A, Nelson TL, Whitfield KE, Kandler C, Jang KL, Gatz M, Butler DA, Stazi MA, Fagnani C, D'Ippolito C, Duncan GE, Buchwald D, Derom CA, Vlietinck RF, Loos RJ, Martin NG, Medland SE, Montgomery GW, Jeong HU, Swan GE, Krasnow R, Magnusson PK, Pedersen NL, Dahl-Aslan AK, McAdams TA, Eley TC, Gregory AM, Tynelius P, Baker LA, Tuvblad C, Bayasgalan G, Narandalai D, Lichtenstein P, Spector TD, Mangino M, Lachance G, Bartels M, van Beijsterveldt TC, Willemsen G, Burt SA, Klump KL, Harris JR, Brandt I, Nilsen TS, Krueger RF, McGue M, Pahlen S, Corley RP, Hjelmberg JV, Goldberg JH, Iwata Y, Watanabe M, Honda C, Inui F, Rasmussen F, Huibregtse BM, Boomsma DI, Sørensen TI, Kaprio J, Silventoinen K. Genetic and environmental influences on adult human height across birth cohorts from 1886 to 1994. *eLife*. 2016;**5**: e20320.
 143. Bechtold S, Beyerlein A, Bonfig W, Dalla Pozza R, Putzker S, Otto R, Schmidt H, Schwarz HP. Sexual difference in bone geometry of adult patients with classical congenital adrenal hyperplasia: data using peripheral quantitative computed tomography. *Horm Res Paediatr*. 2014;**82**(3): 171–178.
 144. Falhammar H, Filipsson H, Holmdahl G, Janson PO, Nordenskjöld A, Hagenfeldt K, Thorén M. Fractures and bone mineral density in adult women with 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2007;**92**(12):4643–4649.
 145. Bhasin S, Storer TW, Berman N, Callegari C, Clevenger B, Phillips J, Bunnell TJ, Tricker R, Shirazi A, Casaburi R. The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. *N Engl J Med*. 1996;**335**(1):1–7.
 146. Moreira CA, Bilezikian JP. Stress fractures: concepts and therapeutics. *J Clin Endocrinol Metab*. 2017; **102**(2):525–534.
 147. Forst-Ludwig A, Kintscher U. Sex differences in exercise-induced cardiac hypertrophy. *Pflugers Arch*. 2013;**465**(5):731–737.
 148. Gibala MJ, Gillen JB, Percival ME. Physiological and health-related adaptations to low-volume interval training: influences of nutrition and sex. *Sports Med*. 2014;**44**(Suppl 2):S127–S137.
 149. Townsend EA, Miller VM, Prakash YS. Sex differences and sex steroids in lung health and disease. *Endocr Rev*. 2012;**33**(1):1–47.
 150. Levine SC, Foley A, Lourenco S, Ehrlich S, Ratliff K. Sex differences in spatial cognition: advancing the conversation. *Wiley Interdiscip Rev Cogn Sci*. 2016; **7**(2):127–155.
 151. Hines M. Prenatal testosterone and gender-related behaviour. *Eur J Endocrinol*. 2006;**155**(Suppl 1): S115–S121.
 152. Hines M, Spencer D, Kung KT, Browne WV, Constantinescu M, Noorderhaven RM. The early postnatal period, mini-puberty, provides a window on the role of testosterone in human neuro-behavioural development. *Curr Opin Neurobiol*. 2016;**38**:69–73.
 153. Pope HG Jr, Kouri EM, Hudson JL. Effects of supraphysiologic doses of testosterone on mood and aggression in normal men: a randomized controlled trial. *Arch Gen Psychiatry*. 2000;**57**(2): 133–140.
 154. Ferguson-Smith MA, Bavington LD. Natural selection for genetic variants in sport: the role of Y chromosome genes in elite female athletes with 46, XY DSD. *Sports Med*. 2014;**44**(12):1629–1634.
 155. Heymsfield SB, Gonzalez MC, Lu J, Jia G, Zheng J. Skeletal muscle mass and quality: evolution of modern measurement concepts in the context of sarcopenia. *Proc Nutr Soc*. 2015;**74**(4):355–366.
 156. Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, Dunkel L, De Lange M, Harris JR, Hjelmberg JV, Luciano M, Martin NG, Mortensen J, Nistico L, Pedersen NL, Skytthe A, Spector TD, Stazi MA, Willemsen G, Kaprio J. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res*. 2003;**6**(5):399–408.
 157. Beunen G, Thomis M. Gene powered? Where to go from heritability (h²) in muscle strength and power? *Exerc Sport Sci Rev*. 2004;**32**(4):148–154.
 158. Silventoinen K, Magnusson PK, Tynelius P, Kaprio J, Rasmussen F. Heritability of body size and muscle strength in young adulthood: a study of one million Swedish men. *Genet Epidemiol*. 2008;**32**(4):341–349.
 159. Seeman E. Pathogenesis of bone fragility in women and men. *Lancet*. 2002;**359**(9320):1841–1850.
 160. Nishiyama KK, Macdonald HM, Moore SA, Fung T, Boyd SK, McKay HA. Cortical porosity is higher in boys compared with girls at the distal radius and distal tibia during pubertal growth: an HR-pQCT study. *J Bone Miner Res*. 2012;**27**(2):273–282.
 161. Oliveira CS, Alves C. The role of the SHOX gene in the pathophysiology of Turner syndrome. *Endocrinol Nutr*. 2011;**58**(8):433–442.
 162. Ottesen AM, Aksglaede L, Garn I, Tartaglia N, Tassone F, Gravholt CH, Bojesen A, Sørensen K, Jørgensen N, Rajpert-De Meyts E, Gerdes T, Lind AM, Kjaergaard S, Juul A. Increased number of sex

- chromosomes affects height in a nonlinear fashion: a study of 305 patients with sex chromosome aneuploidy. *Am J Med Genet A*. 2010;**152A**(5): 1206–1212.
163. Wideman L, Weltman JY, Shah N, Story S, Veldhuis JD, Weltman A. Effects of gender on exercise-induced growth hormone release. *J Appl Physiol*. 1999;**87**(3):1154–1162.
 164. Veldhuis JD, Roemmich JN, Rogol AD. Gender and sexual maturation-dependent contrasts in the neuroregulation of growth hormone secretion in prepubertal and late adolescent males and females—a general clinical research center-based study. *J Clin Endocrinol Metab*. 2000;**85**(7):2385–2394.
 165. Veldhuis JD. Gender differences in secretory activity of the human somatotrophic (growth hormone) axis. *Eur J Endocrinol*. 1996;**134**(3):287–295.
 166. Ho KY, Evans WS, Blizzard RM, Veldhuis JD, Merriam GR, Samojlik E, Furlanetto R, Rogol AD, Kaiser DL, Thorner MO. Effects of sex and age on the 24-hour profile of growth hormone secretion in man: importance of endogenous estradiol concentrations. *J Clin Endocrinol Metab*. 1987;**64**(1):51–58.
 167. Veldhuis JD, Roelfsema F, Keenan DM, Pincus S. Gender, age, body mass index, and IGF-I individually and jointly determine distinct GH dynamics: analyses in one hundred healthy adults. *J Clin Endocrinol Metab*. 2011;**96**(1):115–121.
 168. Veldhuis JD, Patrie JT, Brill KT, Weltman JY, Mueller EE, Bowers CY, Weltman A. Contributions of gender and systemic estradiol and testosterone concentrations to maximal secretagogue drive of burst-like growth hormone secretion in healthy middle-aged and older adults. *J Clin Endocrinol Metab*. 2004;**89**(12):6291–6296.
 169. Roelfsema F, Veldhuis JD. Growth hormone dynamics in healthy adults are related to age and sex and strongly dependent on body mass index. *Neuroendocrinology*. 2016;**103**(3-4):335–344.
 170. Pritzlaff-Roy CJ, Wideman L, Weltman JY, Abbott R, Gutgesell M, Hartman ML, Veldhuis JD, Weltman A. Gender governs the relationship between exercise intensity and growth hormone release in young adults. *J Appl Physiol*. 2002;**92**(5):2053–2060.
 171. Leung KC, Doyle N, Ballesteros M, Sjogren K, Watts CK, Low TH, Leong GM, Ross RJ, Ho KK. Estrogen inhibits GH signaling by suppressing GH-induced JAK2 phosphorylation, an effect mediated by SOCS-2. *Proc Natl Acad Sci USA*. 2003;**100**(3):1016–1021.
 172. Ho KK, O'Sullivan AJ, Wolthers T, Leung KC. Metabolic effects of oestrogens: impact of the route of administration. *Ann Endocrinol (Paris)*. 2003;**64**(2):170–177.
 173. Cappola AR, Bandeen-Roche K, Wand GS, Volpato S, Fried LP. Association of IGF-I levels with muscle strength and mobility in older women. *J Clin Endocrinol Metab*. 2001;**86**(9):4139–4146.
 174. Meinhardt U, Nelson AE, Hansen JL, Birzniece V, Clifford D, Leung KC, Graham K, Ho KK. The effects of growth hormone on body composition and physical performance in recreational athletes: a randomized trial. *Ann Intern Med*. 2010;**152**(9): 568–577.
 175. Harper J. Race times for transgender athletes. *Journal of Sporting Cultures and Identities*. 2015;**6**(1):1–9.
 176. Berman S. Androgens and athletic performance of elite female athletes. *Curr Opin Endocrinol Diabetes Obes*. 2017;**24**(3):246–251.
 177. Elbers JM, Asscheman H, Seidell JC, Megens JA, Gooren LJ. Long-term testosterone administration increases visceral fat in female to male transsexuals. *J Clin Endocrinol Metab*. 1997;**82**(7):2044–2047.
 178. Handelsman DJ. Clinical review: the rationale for banning human chorionic gonadotropin and estrogen blockers in sport. *J Clin Endocrinol Metab*. 2006;**91**(5):1646–1653.
 179. Asbell SO, Leon SA, Tester WJ, Brereton HD, Ago CT, Rotman M. Development of anemia and recovery in prostate cancer patients treated with combined androgen blockade and radiotherapy. *Prostate*. 1996;**29**(4):243–248.
 180. Strum SB, McDermid JE, Scholz MC, Johnson H, Tisman G. Anaemia associated with androgen deprivation in patients with prostate cancer receiving combined hormone blockade. *Br J Urol*. 1997;**79**(6):933–941.
 181. Bogdanos J, Karamanolakis D, Milathianakis C, Repousis P, Tsintavis A, Koutsilieris M. Combined androgen blockade-induced anemia in prostate cancer patients without bone involvement. *Anticancer Res*. 2003;**23**(2C):1757–1762.
 182. Choo R, Chander S, Danjoux C, Morton G, Pearce A, Deboer G, Szumacher E, Loblaw A, Cheung P, Woo T. How are hemoglobin levels affected by androgen deprivation in non-metastatic prostate cancer patients? *Can J Urol*. 2005;**12**(1):2547–2552.
 183. Chander S, Choo R, Danjoux C, Morton G, Pearce A, Deboer G, Szumacher E, Loblaw A, Cheung P, Woo T. Effect of androgen suppression on hemoglobin in prostate cancer patients undergoing salvage radiotherapy plus 2-year buserelin acetate for rising PSA after surgery. *Int J Radiat Oncol Biol Phys*. 2005;**62**(3):719–724.
 184. Golfam M, Samant R, Eapen L, Malone S. Effects of radiation and total androgen blockade on serum hemoglobin, testosterone, and erythropoietin in patients with localized prostate cancer. *Curr Oncol*. 2012;**19**(4):e258–e263.
 185. Storer TW, Miciek R, Travison TG. Muscle function, physical performance and body composition changes in men with prostate cancer undergoing androgen deprivation therapy. *Asian J Androl*. 2012;**14**(2):204–221.

Acknowledgments

The authors are grateful for helpful insights and comments from Alan Verne and Osquel Barroso (World Anti-Doping Agency), Peter Harcourt (Australian Football League, Federation of International Basketball Associations), and Richard Budgett (IOC).

Correspondence and Reprint Requests: David J. Handelsman, PhD, ANZAC Research Institute, University of Sydney, Hospital Road, Concord Hospital, Sydney, New South Wales 2139, Australia. E-mail: djh@anzac.edu.au.

Disclosure Summary: D.J.H. is a medical and scientific consultant for the IAAF and to the Australian Sports Anti-Doping Agency. He is a member of the World Anti-Doping Agency's Health, Medicine and Research Committee and of the IOC working group on hyperandrogenic female and transgender athletes. He has received institutional grant support from Besins Healthcare and Lawley for investigator-initiated clinical studies in testosterone pharmacology and has provided expert testimony in testosterone litigation. A.L.H. is a medical and scientific consultant for the Swedish Olympic Committee and a member of the IAAF and IOC working groups on hyperandrogenic female athletes and transgender athletes. She has received grant support from the IAAF for a study on testosterone and physical performance in women. S.B. is a medical and scientific consultant for the IAAF and a member of the IAAF and IOC working groups on hyperandrogenic female athletes and transgender athletes. The authors have no other involvement with any entity having a financial interest in the material discussed in the manuscript. Opinions expressed in this review are the personal views of the authors and do not represent those of the IAAF, IOC, World Anti-Doping Agency, or Swedish Olympic Committee.

Abbreviations

AR, androgen receptor; CAH, congenital adrenal hyperplasia; CAIS, complete androgen insensitivity syndrome; DSD, disorder (or difference) of sex development; F2M, female-to-male; IAAF, International Association of Athletic Federations; IOC, International Olympic Committee; LC-MS, liquid chromatography–mass spectrometry; M2F, male-to-female; PAIS, partial androgen insensitivity syndrome; PCOS, polycystic ovary syndrome; SHOX, short stature homeobox.

ORIGINAL ARTICLE

WILEY

Sex differences in athletic performance emerge coinciding with the onset of male puberty

David J Handelsman 

ANZAC Research Institute, University of
Sydney, Sydney, NSW, Australia

Correspondence

David J. Handelsman,
ANZAC Research Institute, Concord Hospital 2139
Sydney, NSW, Australia.
Email: djh@anzac.edu.au

Summary

Background: Male performance in athletic events begins to exceed that of age-matched females during early adolescence, but the timing of this divergence relative to the onset of male puberty and the rise in circulating testosterone remains poorly defined.

Design: This study is a secondary quantitative analysis of four published sources which aimed to define the timing of the gender divergence in athletic performance and relating it to the rise in circulating testosterone due to male puberty.

Data: Four data sources reflecting elite swimming and running and jumping track and field events as well as hand-grip strength in nonathletes were analysed to define the age-specific gender differences through adolescence and their relationship to the rising circulating testosterone during male puberty.

Results: The onset and tempo of gender divergence were very similar for swimming, running and jumping events as well as the hand-grip strength in nonathletes, and all closely paralleled the rise in circulating testosterone in adolescent boys.

Conclusions: The gender divergence in athletic performance begins at the age of 12–13 years and reaches adult plateau in the late teenage years with the timing and tempo closely parallel to the rise in circulating testosterone in boys during puberty.

KEYWORDS

age group, performance, puberty, swimming, testosterone, track and field

1 | INTRODUCTION

It is well known that men's athletic performance exceeds that of women especially in power sports because of men's greater strength, speed and endurance. This biological physical advantage of mature males forms the basis for gender segregation in many competitive sports to allow females a realistic chance of winning events. This physical advantage in performance arises during early adolescence when male puberty commences after which men acquire larger muscle mass and greater strength, larger and stronger bones, higher circulating haemoglobin as well as mental and/or psychological differences. After completion of male puberty, circulating testosterone levels in men are consistently 10–15 times higher than in children or women at any age.¹ The age at which sex differences emerge is reported as around the age of 12 from a study of individual Norwegian athletes in two running and two jumping events² and at 13–14 years in four track and field skills in Polish athletes³; however, the

relationship to male puberty and circulating testosterone is not clear. This study investigates the age of the gender divergence in performance in elite swimming and a wider range of elite athletic events as well as a community-based study of grip strength among nonathletes to deduce the onset and progression of the gender divergence in performance of athletes and relates this to the timing and tempo of male puberty and the rise in circulating testosterone into adult male levels.

2 | MATERIAL AND METHODS

Four sources of published data were used in this study for which no ethics approval was required. The first was the US Age Group Swimming time standards which lists the prevailing time standard for entry to the top level (AAAA long course criteria) of all boys and girls events for individual years from 1981 to 2016 (accessed Oct 2016).

DEFENDANT'S
EXHIBIT

29

<http://www.usaswimming.org/DesktopDefault.aspx?TabId=2628&Alias=Rainbow&Lang=en>

Age groups were classified into five categories 10 and under, 11-12 years, 13-14 years, 15-16 years and 17-18 years. The seven events in common to all age groups were freestyle (50 m, 100 m, 200 m), backstroke, breaststroke and butterfly (all 100 m) and individual medley (200 m).

A second data source was the current world records for boys and girls between the ages of 5 and 19 years available at <http://age-records.125mb.com/> (curated by Dominique Eisold, accessed Oct 2016). This included sufficient data to cover the timing of puberty onset with some pre- and postpuberty ages (ages 9-19 years) for a wide range of boys and girls track and field events. For this study, the running events included were 50 m, 60 m, 100 m, 200 m, 300 m, 400 m, 500 m, 600 m, 800 m, 1000 m, 1500 m, 1 mile, 2000 m, 3000 m and 2 miles. Only records recorded by fully automatic timing devices were included whether set indoor or outdoor or at altitude (>1000 m), but wind-assisted records were excluded from this analysis. The jumping events included were high jump, pole vault, long jump, triple jump, standing long jump.

The third data source was from a published study¹ in which serum testosterone was measured in over 100 000 consecutive serum samples processed over 7 years from a single pathology laboratory which was analysed to estimate male and female age-specific reference ranges across the full lifespan.

The fourth was a meta-analysis of secular changes in hand-grip strength in nonathletic children and adolescents from Canada and United States⁴ using the data provided on 5676 males and 5489 females in 19 studies conducted between 1966 and 2009.

Data analysis was performed by analysis of variance and nonlinear curve fitting using NCSS 11 Statistical Software (NCSS LLC, Kaysville, Utah, USA). For each event used in this analysis, the age-specific record or age-group time standard was defined for boys (Tb) and girls (Tg) so the difference (expressed as a percentage) between boys and girls for any event was defined as $D = (Tg - Tb) * 100 / Tg$. For athletic jumping events, an analogous definition for record length was used

(Lb for boys, Lg for girls) with the male advantage defined as $D = (Lb - Lg) * 100 / Lg$. For the athletic events where individual year age records were available across the age of puberty, the age-specific difference (as a percentage) for each year of age were pooled into running or jumping categories. For track and field performance, the pooled data were fitted to a four-parameter sigmoidal curve which allowed for asymptotic estimation of the lower (prepubertal) and upper (postpubertal) plateaus from the four parameters. In addition, the timing and tempo of the pubertal increase were defined by the start of puberty, defined as the time when 20% of the ultimate increase due to puberty had occurred (ED_{20}), and mid-puberty as the time when half the ultimate increase had occurred (ED_{50}). For swimming, the pooled gender differences for all strokes and distances were fitted by a smoothed spline curve. For hand-grip strength, the differences were fitted to a piecewise linear-quadratic curve with a single inflexion point.

3 | RESULTS

In swimming performance, the overall gender differences were highly significant with age group ($F_{4,360} = 1481$, $P < .0001$) and stroke ($F_{4,360} = 11.9$, $P < .0001$) as main (between) effects (Figure 1). There was no significant difference according to year (as a within factor, $P = .99$) so that for further analysis, years were taken as replicates. Using a sigmoidal curve fit for the overall gender differences pooling all strokes and distances, the ED_{20} was 11.4 years and the ED_{50} was 12.8 years.

Within a single stroke (freestyle), in addition to expected age-group effects ($F_{4,525} = 2174$, $P < .0001$), there were also significant effects according to distance ($F_{2,525} = 231.5$, $P < .0001$) whereby the age-group effects was significantly greater the shorter the event distance (Figure 2, 50 m > 100 m > 200 m, age group x distance interaction, $F_{8,525} = 55.9$, $P < .0001$) (Figure 1). Similarly, for a fixed length of events (100 m) and after taking age-group effects into account, the four form strokes did differ significantly ($F_{3,700} = 12.9$, $P < .0001$) producing significant

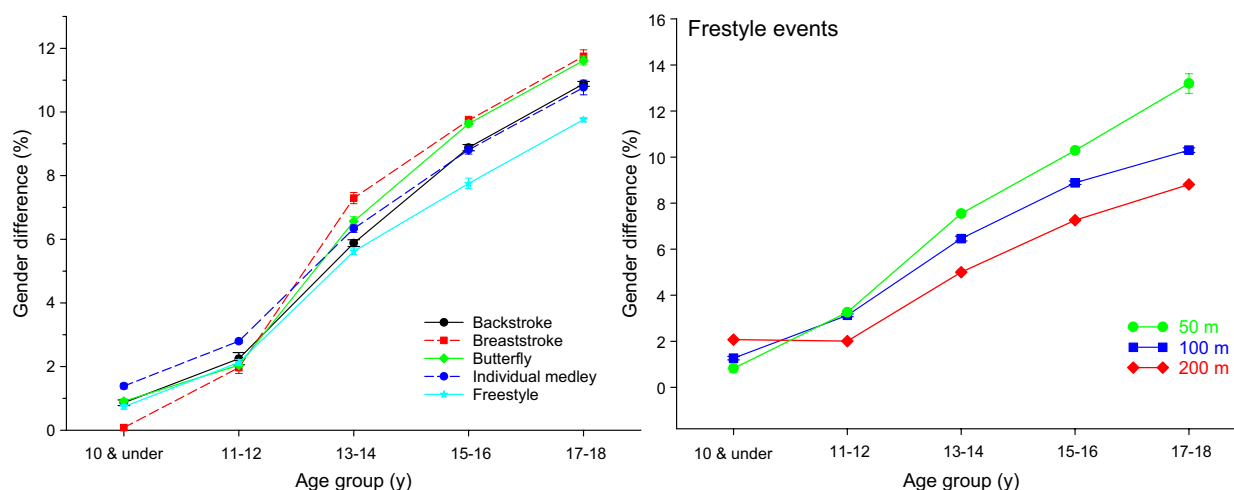


FIGURE 1 Gender differences in performance (in percentage) according to age group and stroke (left panel) or distance in freestyle events (right panel) in swimming events. Data shown as mean and standard error of the mean. Note greatest increase after the age of 12 years by age in breaststroke and least in freestyle and magnitude of increases are 50 m > 100 m > 200 m in freestyle events. [Colour figure can be viewed at wileyonlinelibrary.com]

differences between strokes (interaction $F_{12,700}=23.4$, $P<.0001$), the most prominent being for breaststroke, which displayed the greatest age-group effect, and butterfly followed by backstroke and then free-style, which showed the least age-group effect (Figure 1).

In track and field athletics, the effects of age on running performance (Figure 2 upper left panel) showed that the prepubertal differences of 3.0% increased to a plateau of 10.1% with an onset (ED_{20}) at 12.4 years and reaching midway (ED_{50}) at 13.9 years. For jumping (Figure 2 upper right panel), the prepubertal difference of 5.8% increased to 19.4% starting at 12.4 years and reaching midway at 13.9 years. The timing of the male advantage in running, jumping and swimming was similar and corresponded to the increases in serum testosterone in males (Figure 2 lower panel).

To examine age of gender divergence in strength in an analogous data set from a nonathletic population (Canadian and US children and adolescents), the age trends in hand-grip strength showed a difference in hand-grip strength commencing from the age of 12.8 years onwards (Figure 3). Prior to the age of 13 years, boys had a marginally significant greater grip strength than girls ($n=45$, $t=2.0$, $P=.026$), but after the

age of 13 years, there was a strong significant relationship between age and difference in grip strength ($n=18$, $r=.89$, $P<.001$).

4 | DISCUSSION

The present study shows that the gender divergence in performance for swimming and for running and jumping track and field events is very closely aligned to the timing of the onset of male puberty, which typically has onset at around 12 years of age.^{5,6} These findings are consistent with reports on the timing of the gender differences in performance observed among Norwegian athletes in two running and two jumping events² and for track and field skills among Polish athletes.³ This study extends the findings to swimming and a wider range of running and jumping track and field events. This timing is also consistent with the start of the gender divergence in fat-free (muscle) mass⁷ and strength increases.^{8,9}

In this study, the timing and tempo of male puberty effects on running and jumping performance were virtually identical and very similar

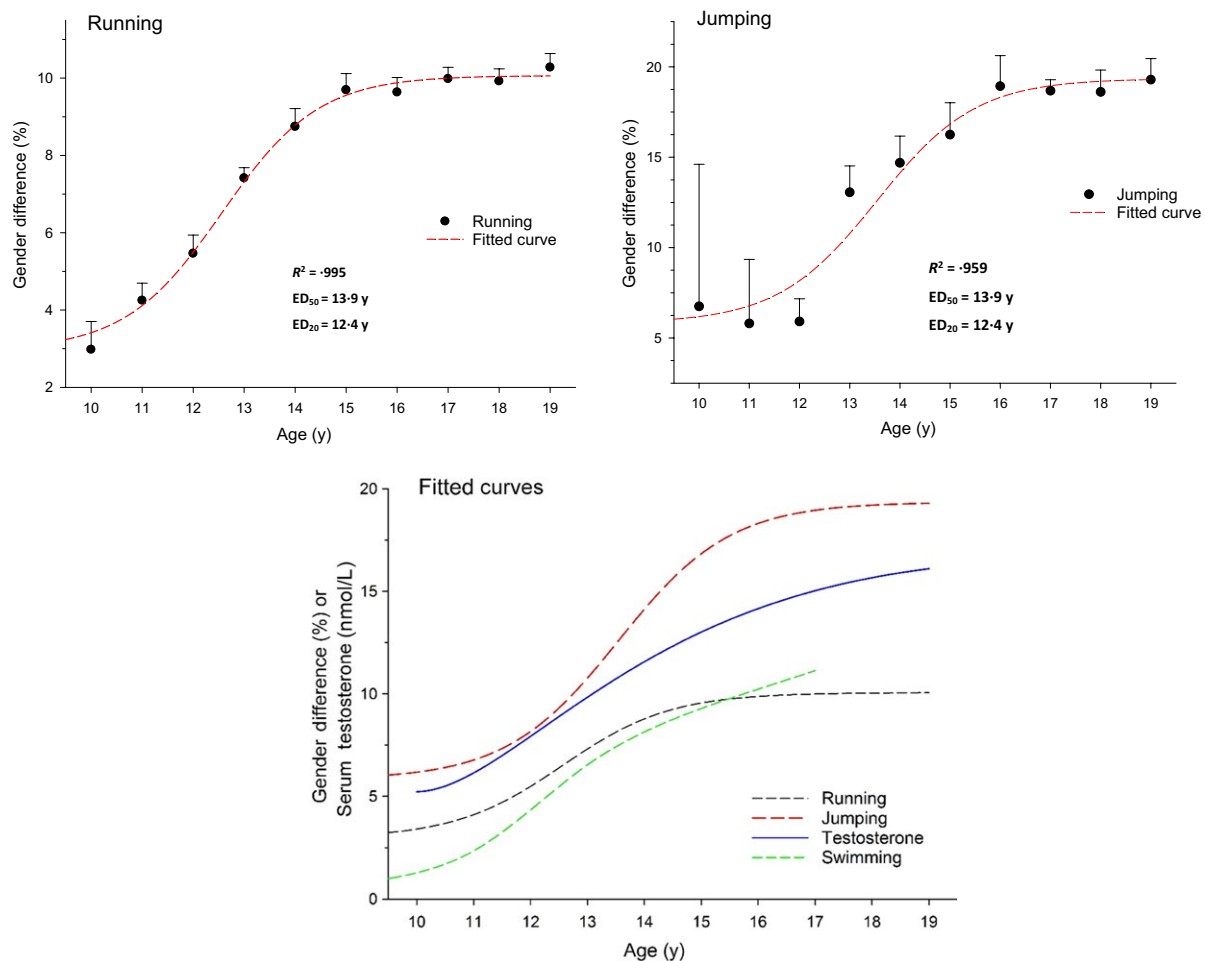


FIGURE 2 Gender differences in performance (in percentage) according to age (in years) in running events including 50 m, 60 m, 100 m, 200 m, 300 m, 400 m, 500 m, 600 m, 800 m, 1000 m, 1500 m, 1 mile, 2000 m, 3000 m and 2 miles (upper left panel) and in jumping events including high jump, pole vault, triple jump, long jump and standing long jump (upper right panel). Fitted sigmoidal curve plot of gender differences in performance (in percentage) according to age (in years) in running, jumping and swimming events as well as serum testosterone (lower panel). Data shown as mean and standard error of the mean of the pooled gender differences by age. [Colour figure can be viewed at wileyonlinelibrary.com]

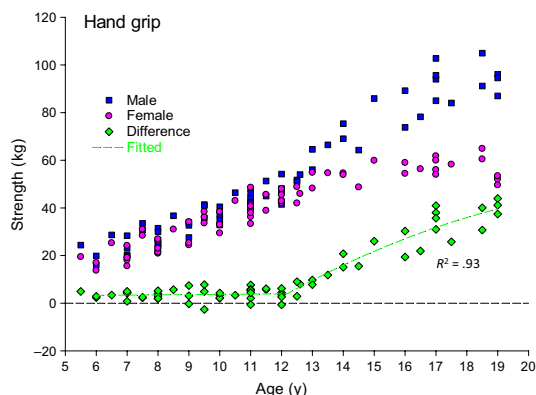


FIGURE 3 Hand-grip strength in children and adolescents from 19 studies including 5676 males (square) and 5489 females (circles) and the differences between male and females (diamonds) conducted between 1966 and 2009. The dotted line represents the fitted curve using a piecewise linear-quadratic curve fit with an automatically defined inflexion point at 12.8 years. [Colour figure can be viewed at wileyonlinelibrary.com]

to those in swimming events. Furthermore, these coincided with the timing of the rise in circulating testosterone due to male puberty. In addition to the strikingly similar timing and tempo, the magnitude of the effects on performance by the end of this study was 10.0% for running and 19.3% for jumping, both consistent with the gender differences in performance of adult athletes previously reported to be 10%–12% for running^{10,11,12} and 19% for jumping.¹² The similar magnitude of the plateau effects observed for the oldest (postpubertal) stages in this study with mature adult gender differences suggests there are likely minimal if any further divergences in gender performance among athletes after the age of 20 years.

In the swimming events, despite the continued progressive improvements in individual male and female event records, the stability of the gender difference over 35 years shown in this study suggests that the gender differences in performance are stable and robust. These findings are consistent with a previous report of no narrowing of the gender gap in swimming event performance over more than three decades.¹² These findings contribute to discounting previous suggestions that the gender gap in performance of athletes was narrowing and might even disappear,¹³ interpretations which were confounded by the increasing participation of females in elite sports through the 20th century that led to short-term accelerating improvement until women approached closer to contemporary female performance plateau.¹² The greater effect of male puberty on shorter freestyle events is consistent with the greater power demands of short sprint events than for longer freestyle events that involve more endurance. The consistent differences between form strokes over 100-m events, even after accounting for the very dominant age-group effect, suggest that the power demands on performance were most prominent in breaststroke and least in freestyle, presumably due to the different mechanical demands of the different strokes.

The gender divergence in hand-grip strength among nonathletic children and adolescents strengthens the view that these gender divergences are a feature of normal male puberty rather than being a feature that manifests only in elite athletes.

The similar time course of the rise in circulating testosterone with that of the gender divergences in swimming and track and field sports is strongly suggestive that these effects arise from the increase in circulating testosterone from the start of male puberty.¹ Somatic effects of male puberty differ in responsiveness to the postpubertal increase in serum testosterone. Muscle effects of testosterone have been established in well-controlled, interventional clinical experiments in healthy young^{14,15} and older¹⁶ men. Testosterone increases muscle mass and strength over weeks to months with a strong dose-response evident from below to above physiological testosterone doses and concentrations. Analogous findings are reported in androgen-deficient (hypogonadal) men administered testosterone replacement therapy¹⁷ and in women receiving appropriately lower testosterone doses,¹⁸ and observational dose-effect relationship between endogenous testosterone and upper or lower body muscle mass is reported in healthy men.¹⁹ Most if not all sex differences in maximal oxygen uptake are explained by differences in muscle mass.^{20–22}

Adult male circulating testosterone also has marked effects on bone development leading to longer, stronger and denser bone than in age-matched females.²³ However, testosterone effects on bone are slower in onset and probably less reversible than effects on muscle. For example, men achieve peak bone mass at the end of skeletal maturation only in the early 1920s, about a decade after the start of sustained exposure to adult male testosterone levels. Furthermore, while testosterone deficiency may lead to loss of bone density,²³ the overall structural framework of the skeleton is likely to change slowly if at all. Hence, the extent to which testosterone-induced bone changes contribute to the male advantage in adolescent athletic performance is unclear but is probably at least not maximal until the third decade of life by which time the gender differences are already stabilized.

A further biological advantage of adult male circulating testosterone concentrations is the increased circulating haemoglobin. Men have ~10 g/L greater haemoglobin than women²⁴ with the gender differences also evident from the age of 13–14 years.²⁵ Testosterone effects on haemoglobin are replicated by administration of exogenous testosterone in a dose-dependent fashion²⁶ within 1–3 months.²⁷ Like the effects on muscle, the erythropoietic effect of testosterone is relatively rapid and reversible in contrast to the slower effects on bone. Although a higher haemoglobin is likely to provide advantages in endurance rather than power events, it is unclear how much the relatively modest magnitude of this gender difference contributes to the male advantage in athletic performance.

Finally, exposure to adult male testosterone concentrations is likely to produce some mental or psychological effects.²⁸ However, the precise nature of these remains controversial and it is not clear whether, or to what extent, this contributes to the superior elite sporting performance of men in power sports compared with the predominant effects on muscle mass and function.

The strength of the present study is that it includes a wide range of swimming as well as track and field running and jumping events as well as strength for nonathletes for males and females across the ages spanning the onset of male puberty. The similar timing of the gender divergence in each of these settings to that of the rise in circulating

testosterone to adult male levels strongly suggests that they all reflect the increase in muscular size and strength although the impact of other androgen-dependent effects on bone, haemoglobin and psychology may also contribute. Limitations of this study include that it could not extend to all swimming or track and field events due to the restricted participation of younger age groups in more gruelling events. Furthermore, the testosterone measurements were not from the individual athletes included in the analysis of available published data so that the comparisons are cohort-wise rather than based on individuals.

It is concluded that the gender divergence in athletic performance begins at the age of 12-13 years and reaches adult plateau in the late teenage years. Although the magnitude of the divergence varies between athletic skills, the timing and tempo are closely parallel with each other and with the rise in circulating testosterone in boys during puberty to reach adult male levels.

ACKNOWLEDGEMENTS

The author is grateful to Professor Ken Fitch for helpful advice.

CONFLICT OF INTERESTS

Nothing to declare.

REFERENCES

- Handelsman DJ, Sikaris K, Ly LP. Estimating age-specific trends in circulating testosterone and sex hormone-binding globulin in males and females across the lifespan. *Annu Clin Biochem*. 2016;53:377-384.
- Tonnessen E, Svendsen IS, Olsen IC, Guttormsen A, Haugen T. Performance development in adolescent track and field athletes according to age, sex and sport discipline. *PLoS ONE*. 2015;10:e0129014.
- Malina RM, Slawinska T, Ignasiak Z, et al. Sex differences in growth and performance of track and field athletes 11-15 years. *J Hum Kinet*. 2010;24:79-85.
- Silverman IW. The secular trend for grip strength in Canada and the United States. *J Sports Sci*. 2011;29:599-606.
- Beccuti G, Ghizzoni L. Normal and abnormal puberty. In: De Groot LJ, Beck-Peccoz P, Chrousos G, et al., eds. *Endotext*. MDText.com, Inc.: South Dartmouth, MA; 2000.
- Day FR, Bulik-Sullivan B, Hinds DA, et al. Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nat Commun*. 2015;6:8842.
- Malina RM, Bouchard C, Beunen G. Human growth: selected aspects of current research on well-nourished children. *Ann Rev Anthropol*. 1988;17:187-219.
- Sartorio A, Lafortuna CL, Pogliaghi S, Trecate L. The impact of gender, body dimension and body composition on hand-grip strength in healthy children. *J Endocrinol Invest*. 2002;25:431-435.
- Henneberg M, Brush G, Harrison GA. Growth of specific muscle strength between 6 and 18 years in contrasting socioeconomic conditions. *Am J Phys Anthropol*. 2001;115:62-70.
- Cheuvront SN, Carter R, Deruisseau KC, Moffatt RJ. Running performance differences between men and women: an update. *Sports Med*. 2005;35:1017-1024.
- Seiler S, De Koning JJ, Foster C. The fall and rise of the gender difference in elite anaerobic performance 1952-2006. *Med Sci Sports Exerc*. 2007;39:534-540.
- Thibault V, Guillaume M, Berthelot G, et al. Women and men in sport performance: the gender gap has not evolved since 1983. *J Sports Sci Med*. 2010;9:214-223.
- Beneke R, Leithauser RM, Doppelmayer M. Women will do it in the long run. *Br J Sports Med*. 2005;39:410.
- Bhasin S, Storer TW, Berman N, et al. The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. *N Engl J Med*. 1996;335:1-7.
- Finkelstein JS, Lee H, Burnett-Bowie SA, et al. Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med*. 2013;369:1011-1022.
- Bhasin S, Woodhouse L, Casaburi R, et al. Older men are as responsive as young men to the anabolic effects of graded doses of testosterone on the skeletal muscle. *J Clin Endocrinol Metab*. 2005;90:678-688.
- Bhasin S, Storer TW, Berman N, et al. Testosterone replacement increases fat-free mass and muscle size in hypogonadal men. *J Clin Endocrinol Metab*. 1997;82:407-413.
- Huang G, Basaria S, Travison TG, et al. Testosterone dose-response relationships in hysterectomized women with or without oophorectomy: effects on sexual function, body composition, muscle performance and physical function in a randomized trial. *Menopause*. 2014;21:612-623.
- Mouser JG, Loprinzi PD, Loenneke JP. The association between physiologic testosterone levels, lean mass, and fat mass in a nationally representative sample of men in the United States. *Steroids*. 2016;115:62-66.
- Jones NL, Makrides L, Hitchcock C, Chypchar T, McCartney N. Normal standards for an incremental progressive cycle ergometer test. *Am Rev Respir Dis*. 1985;131:700-708.
- Svedenhag J. Maximal and submaximal oxygen uptake during running: how should body mass be accounted for? *Scand J Med Sci Sports*. 1995;5:175-180.
- Genberg M, Andren B, Lind L, Hedenstrom H, Malinowski A. Commonly used reference values underestimate oxygen uptake in healthy, 50-year-old Swedish women. *Clin Physiol Funct Imaging*. 2016; doi: 10.1111/cpf.12377. [Epub ahead of print]
- Vanderschueren D, Laurent MR, Claessens F, et al. Sex steroid actions in male bone. *Endocr Rev*. 2014;35:906-960.
- Murphy WG. The sex difference in haemoglobin levels in adults - mechanisms, causes, and consequences. *Blood Rev*. 2014;28:41-47.
- Krabbe S, Christensen T, Worm J, Christiansen C, Transbol I. Relationship between haemoglobin and serum testosterone in normal children and adolescents and in boys with delayed puberty. *Acta Paediatr Scand*. 1978;67:655-658.
- Coviello AD, Kaplan B, Lakshman KM, Chen T, Singh AB, Bhasin S. Effects of graded doses of testosterone on erythropoiesis in healthy young and older men. *J Clin Endocrinol Metab*. 2008;93:914-919.
- Bachman E, Travison TG, Basaria S, et al. Testosterone induces erythrocytosis via increased erythropoietin and suppressed hepcidin: evidence for a new erythropoietin/hemoglobin set point. *J Gerontol A Biol Sci Med Sci*. 2014;69:725-735.
- Celec P, Ostatnikova D, Hodosy J. On the effects of testosterone on brain behavioral functions. *Front Neurosci*. 2015;9:12.

How to cite this article: Handelsman DJ. Sex differences in athletic performance emerge coinciding with the onset of male puberty. *Clin Endocrinol (Oxf)*. 2017;87:68-72. <https://doi.org/10.1111/cen.13350>

Open camera or QR reader and
scan code to access this article
and other resources online.



Pediatric Provider Perspectives on Laws and Policies Impacting Sports Participation for Transgender Youth

Landon D. Hughes,^{1,2} Nadia Dowshen, MD,^{3,4} Kacie M. Kidd, MD,⁵ Don Operario, PhD,⁶
Christopher Renjilian, MD,^{3,7} and Kristi E. Gamarel, PhD^{1,2}

Abstract

Purpose: Since May 2021, numerous state legislatures in the United States have introduced bills to restrict the participation of transgender and gender diverse (trans) youth in gender-segregated sports in a manner consistent with their gender identity. The purpose of this study was to understand how medical providers of pediatric gender-affirming care for trans youth view these legislative efforts and how they believe these bills will affect their practice and patients.

Methods: In March 2021, we recruited participants using listservs known to be frequented by providers of gender-affirming medical care to complete a survey about bills that restrict trans youths' participation in sex-segregated sports. Eligible participants were over the age of 18, currently worked as a physician, nurse practitioner, or physician's assistant, and provided gender-affirming care to trans youth under the age of 18 in the United States.

Results: We analyzed the responses of 103 providers from all 50 states and the District of Columbia. Most participants identified as White (77%), cisgender women (70%), and specialized in pediatric care (52%). The most salient theme, described by nearly all participants, was that legislation banning trans youth from sports participation would lead to worsening discrimination and stigmatization. Other themes included worsening mental and physical health of trans youth, forced changes to clinical practice, politicization of trans youth, and efforts required to stop these bills.

Conclusions: Providers of gender-affirming care in this study overwhelmingly opposed legislation that bans trans youth from sports participation citing the severe consequences to the well-being of trans youth. More research is needed to examine stakeholder's opinions regarding legislation that does not ban but otherwise restricts sports participation by trans youth.

Keywords: adolescent, health personnel, laws and statutes, sports, sports participation, transgender

Introduction

THE BENEFITS OF organized sports participation to the health of youth are numerous and well-established.¹ Research has shown that organized sports protect youth against cardiovascular disease, obesity, and risk-taking behaviors.^{2–7} The numerous benefits of sports participation have been shown to last into early adulthood; for example, 1 study followed 462 youth who participated in organized sports into early adulthood and found that

sports participation was associated with lower body-mass index, higher high-density lipoprotein, and increased physical activity.⁷

In addition to physical health benefits, participation in sports has been shown to improve the mental health of youth, decreasing the risk of depression as well as social and emotional problems.^{8,9} Compared to their cisgender peers, transgender and gender diverse (trans) youth have been shown to be at increased risk for mental health problems, including depression, anxiety, and suicidal ideation.^{10,11}

¹Department of Health Behavior and Health Education, University of Michigan School of Public Health, Ann Arbor, Michigan, USA.

²Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA.

³Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA.

⁴Division of Adolescent Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.

⁵Department of Pediatrics, West Virginia University School of Medicine, Morgantown, West Virginia, USA.

⁶Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, Rhode Island, USA.

⁷Sports Medicine and Performance Center, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.

DEFENDANT'S
EXHIBIT

30

Among trans youth, sports participation has been associated with increased self-esteem and sense of school belonging and decreased risk of depression.¹² While sports participation offers several mental and physical health benefits, it is important to note that *forced* participation in sports may be particularly harmful for trans youth, especially if the participation is on teams that are not aligned with a person's gender identity or the environment is unwelcoming.^{12,13} Therefore, when we discuss sports participation, we mean *voluntary* sports participation.

Despite the known benefits of organized sports to the long-term health of youth, between January 2020 and May 2021, 34 state legislatures have introduced bills limiting the participation of transgender youth in sex-segregated sports, with most focused on high school sports and preventing youth assigned male at birth from participating on sex-segregated girls' teams.¹⁴ These bills reflect the concern of some that those assigned male at birth are conferred androgen-related advantage in stature and strength compared to those assigned female.

Since May 2021, eight states have passed these bills as law.^{14,15} The majority of these bills prevent trans girls from participating in sports with cisgender girls,^{16–19} while others require that any trans person participates only in gender-segregated sports with those who share their sex assigned at birth (unless as outlined in Texas Senate Bill 29, a designated competition for female athletes is unavailable).^{20,21} These laws stand in opposition to policies instituted by the International Olympic Committee, which does not ban trans athletes from competing in sex-segregated competition solely based on their sex assigned at birth.²²

Before the introduction of these bills, research documented stigma, discrimination, and structural barriers already faced by trans athletes in sports and sports-related physical activity.^{23–25} In several qualitative studies, trans people have reported unwelcoming environments and policies as being among the greatest barriers to sports participation.^{23,24} One study that included over 8000 trans students between the ages of 13 and 21 found that trans students were less likely to participate in sports and more likely to avoid locker rooms than their cisgender gay and lesbian counterparts when school policies required that they use locker rooms and participate on teams associated with their sex assigned at birth instead of their gender identity.¹¹

Medical providers who provide gender-affirming care for trans youth are a key group of experts who can help to inform policy discussions in the public forum. Medical providers work closely with trans youth and their families to provide gender-affirming medical care (e.g., puberty blockers, hormones) and to monitor the physical health and mental well-being of trans youth before and throughout their social and/or medical transition.²⁶ Providers frequently educate the public, school administrators, and sports organizations about how to include trans athletes in competitive sports and ensure supportive environments for trans athletes.^{23,27,28} Providers of gender-affirming care to trans youth are uniquely situated to understand the effects these bills have on the physical and mental health of the trans youth they serve.

Thus, the purpose of this study was to understand how medical providers of gender-affirming care view legislation that restricts trans youth from competition in gender-segregated sports in a manner consistent with their gender identity and how they believe these bills will affect their practice and patients. While it would be valuable to elicit

perspectives from a wide variety of medical providers, we chose to focus on those who provide gender-affirming care because these providers are the best equipped to accurately predict the effects of such bills on trans youth.

Materials and Methods

Recruitment and survey development

In March of 2021, we recruited participants using professional society listservs for providers who practice pediatric gender-affirming medical care to complete a survey about bills that ban trans youth from participating in gender-segregated sports in a manner consistent with their gender identity. We asked individuals to share the survey link through providers' professional networks and clinics. To be eligible for our study, providers needed to (1) be at least 18 years of age, (2) currently work as a physician (MD/DO), nurse practitioner, or physician's assistant, and (3) currently provide gender-affirming medical care (e.g., prescribe puberty blockers or hormones) to trans youth under the age of 18 in the United States.

The survey was only available in English, and data were collected using Qualtrics. Participants' consent for the study was collected electronically. Participants were not compensated. Our study was ruled exempt by the University of Michigan's institutional review board (HUM00196292).

The survey was developed by a multidisciplinary team of researchers and medical providers who are committed to addressing inequities among trans youth. Before fielding the survey, five medical providers with relevant clinical expertise beta-tested the survey to ensure the content was clear and appropriate for the intended audience and that the skip logic worked appropriately.

Survey instrument

Before responding, participants were provided with the following information: "Several state legislatures in the United States are currently considering or have passed laws that would ban transgender and gender diverse athletes from participating in gendered sports (i.e., banning a transgender girl from a girls only team)." Participants were asked to provide their thoughts about these proposed laws in four separate open-ended survey questions: "What do laws like this mean to you as a gender-affirming care provider for transgender and gender diverse youth?" "How do you think laws like this would impact your practice?" "How do you think laws like this would impact your patients?" "What steps, if any, do you think would be helpful to ensure transgender and gender diverse youth are not banned from participating in sports?"

In addition, participants were asked to provide information on their demographics, medical practice, and the types of gender-affirming care they provide. Gender identity was collected by asking participants "Which gender do you identify with now? Select all that apply." The options were listed as follows: man, woman, trans man, trans woman, non-binary or genderqueer, and other please specify. We also asked, "What sex were you assigned at birth on your original birth certificate?" with the options of male and female being listed.

We then coded those who identified as women and assigned female at birth as "cisgender women," those who identified as men and assigned male at birth as "cisgender men," those who identified themselves as trans men as

TABLE 1. PARTICIPANT DEMOGRAPHICS ($N=103$)

	M (SD) n (%)
Age	43 (9.4) n (%)
Gender identity ^a	
Cisgender women	65 (63)
Cisgender men	22 (21)
Trans men	2 (2)
Nonbinary or genderqueer	4 (4)
Race/ethnicity ^a	
Non-Hispanic White	72 (70)
Non-Hispanic Asian	5 (5)
Non-Hispanic Black	2 (2)
Non-Hispanic Middle Eastern, North African, Arab, or Chaldean	2 (2)
Hispanic or Latinx	4 (4)
Biracial or other	7 (7)
Current occupation	
Doctor (MD/DO)	100 (97)
Nurse practitioner	2 (2)
Physician's assistant	1 (1)
Primary training ^b	
Pediatrics	54 (52)
Adolescent medicine	39 (38)
Endocrinology	32 (31)
Family medicine	10 (10)
Other	6 (6)
Census region where providers practice ^b	
Midwest	40 (39)
Northeast	57 (55)
South	51 (50)
West	52 (50)
Gender-affirming care provided ^b	
Prescribed hormones	96 (93)
Prescribed menstrual suppressants	91 (88)
Prescribed antiandrogens	89 (86)
Surgery referrals	88 (85)
Prescribed puberty blockers	83 (81)
Letter writing to support medical care	81 (79)
Mental health care	18 (17)
Time providing gender-affirming care to trans youth ^a	
<1 year	4 (4)
1–3 years	23 (22)
4–6 years	35 (34)
7–9 years	22 (21)
10–19 years	15 (15)
20+ years	3 (3)
Number of patients who are trans youth ^a	
<5	2 (2)
5–10	12 (12)
11–50	29 (28)
51–100	16 (16)
101–150	10 (10)
151–200	8 (8)
201+	25 (24)

^aNote that all responses were optional, therefore some categories do not total to 100%.

^bGreater than total n as participants are able to choose more than one option.

DO, Doctor of Osteopathic Medicine; M , mean; MD, Doctor of Medicine; SD, standard deviation.

“trans men,” and those who identified as nonbinary or genderqueer as “non-binary or genderqueer.” For all other demographic items found in Table 1, the available choices for these items are exactly as they appear in the table.

Data analysis

Using the demographic information provided, we conducted univariate analyses to characterize the study sample. We then used an inductive thematic approach to analyze open-ended responses.²⁹ Two members of the study team familiarized themselves with each response. Then, they identified broad themes and refined a codebook through discussions. The first author then indexed the data and identified specific sections that corresponded to our themes. Discrepancies in interpretation were resolved through discussion with the entire team.

Results

Overall, 240 participants were screened, and 103 eligible providers participated in the study. Table 1 shows the majority of participants identified as cisgender women (63%). Participants ranged in age from 27 to 69 ($M=43$, standard deviation=9.4), and the majority (73%) have been providing gender-affirming care to trans youth for over 4 years. Most participants identified as White (70%). Our participants represented all 50 states, and the District of Columbia, and were spread evenly across Census Regions. The majority of respondents were physicians (97%), followed by nurse practitioners (2%), and physician's assistants (1%). Most participants specialized in pediatrics (52%). Summarized below are the five overarching themes identified from the survey responses.

Theme 1: increased discrimination and stigmatization of trans youth

Nearly every provider believed that these bills would increase the stigmatization of trans youth. Providers discussed that trans youth would be deeply impacted by the sudden exclusion from sports and teams they have been enjoying since their childhood.

“It would be horrible for the kids I see who enjoy sports, particularly those who had a relatively early social transition and are accustomed to full participation in life in their affirmed gender... ‘I can go to school. I can go to church. I can go to my friend's house and shoot hoops....why can't I be on the basketball team? My [other cisgender] friends are.’” (Georgia)

Others described that the mere introduction of these bills increases stigma and shame for trans youth, regardless of their passage.

“[These bills] impact their fear of discrimination, shame around their bodies, likely to reduce their participation in sports even if not banned over the feeling that there is something wrong with their participation.” (California)

Several noted that these bills may increase violence and be a threat to young people's physical safety. They also thought that stigma as a result of these bills might delay their transition or make them less likely to participate in sports in the future.

“I’m sure [these bills] will be devastating for some transgender young athletes who have a passion for their sport. They will be forced to choose between their sport and their gender identity.” (Michigan)
 “[These bills] may instill fear or hesitation in patients/families considering gender-affirming care interventions.” (Washington)

Theme 2: worsen mental and physical health of trans youth

Most providers believed that these laws would increase mental health problems among their patients (e.g., depression, suicidality, self-harm, distress, gender dysphoria, trauma).

“This is another example of ignorance, bigotry, and discrimination. This type of exclusion contributes to social othering and the increased rates of anxiety, suicide, and depression observed among trans youth. Not to mention a transgender girl being forced to play on the boys’ team, especially in states where such laws may be implemented, likely poses a threat to the physical safety of that transgender girl to say nothing of her emotional well-being. So many youth benefit from team sports and to deny that to any child is, in my opinion, unjust.” (Rhode Island)

They also believed that some of these laws¹⁷ may require genital examinations that might be traumatic for some patients. Many providers believed that these laws would worsen the physical health of their patients, particularly increasing diabetes and obesity given a decrease in physical activity.

“[These bills] mean that my patients [would] have less access to healthy activities and they are already at higher risk for unhealthy eating behaviors and weight.” (North Carolina)

Some providers believed that their patients would be less likely to get medical care for fear of being “outed.” Some providers recognized that school requirements for preparticipation sports evaluations can serve as an important opportunity to bring young people into the medical office for checkups and routine preventative health care. When some youth are fully excluded from sports and sports-related requirements on the basis of their gender identity, they may then have fewer opportunities for preventative health care as their peers.

“[Trans youth] would not need sports exams and may miss health issues and immunizations due to not coming in.” (Tennessee)

Theme 3: force changes to medical practices

Providers thought that these laws would change their practice in a myriad of ways. Several providers anticipated the need to increase counseling services or provide more time in their visits to address emotional distress and mental health issues with trans patients.

“I will need to spend time processing the anger, sadness, devastation of youth who wish to participate in sports like other kids.” (New York)

Several providers also described having to identify alternative strategies to encourage their trans patients to get physically active and promote psychological wellness, should these bills pass.

“Sports participation is key to maintaining one’s mental health. Disrupting this will further result in increased harm.” (Maryland)

“I think [these bills] would make it more difficult to provide adequate care by limiting options available for gender diverse athletes.” (Illinois)

Theme 4: politicization of trans youth

Many providers believed that these laws represent a political overreach and limitation to the rights of their patients.

“[These bills] would be devastating for my patients, many of whom are athletes. This is gender-based discrimination and should be recognized as such.” (Pennsylvania)

Many stated that these laws are creating a problem where one does not currently exist. Several believed that these bills are not consistent with the science of hormonal effects on athletic performance and indicated that such policy-making attempts should incorporate expertise from the medical community and sports organizations.

“[These bills show a] lack of understanding and guidance for what stages and hormones levels could change someone’s muscular mass or abilities.” (Oregon)

“[Lawmakers should] allow experts (Pediatric Endocrine Society-PES/World Professional Association for Transgender Health-WPATH, others) to establish participation guidelines based on facts and medical research.” (Tennessee)

Several providers described these bills as “unscientific” because the bills do not consider the mental, physical, and social health benefits of sports participation.

Theme 5: efforts required to stop bills

All providers believed that these bills should be stopped. Most believed the way to do this is through advocacy. Particularly, providers believed that medical associations, along with collegiate and professional athletic associations, should publicly oppose these laws and speak up on behalf of trans youth. When asked about what could be done to ensure trans youth are able to participate in sports, one provider said:

“[H]aving all medical societies write official position papers opposing these potential laws. Also, having medical providers/societies work to advocate nationally would be important.” (New York)

Several providers believed that their professional peers should educate the public on the importance of participation in sports for health and dispel myths about hormonal advantages, particularly among youth. Some providers also noted that legislation ensuring that trans youth are able to participate in sports should be passed. Some providers wanted more research to better understand the issue and inform the public.

Discussion

This study sought to understand how providers of pediatric gender-affirming medical care view bills that would bar trans youth from participating in gender-segregated sports in a manner consistent with their gender identity, an issue that has garnered significant political attention and attempts at policy intervention. Perspectives from medical providers are crucial to the analysis of these policy initiatives, given their professional expertise on the effects of gender-affirming

medical care, the physical and mental health benefits of sports, and their proximity to trans youth and their families. We were particularly interested in how providers thought these bills would affect their practice and patients.

Providers believed that these bills would increase the number of mental and physical health morbidities among trans youth, especially those assigned male at birth. Particularly, providers believed that these bills would increase the risks of suicide, depression, nonsuicidal self-injury, anxiety, and gender dysphoria.^{10,11} For example, a substantial body of literature has shown that trans youth are already at increased risk for mental health conditions, such as suicide and suicidal ideation, than their cisgender peers, which could be exacerbated by the introduction of these bills, with several studies suggesting that these bills could even worsen the mental health of those who aren't directly targeted by these laws (e.g., trans youth who don't play sports).^{30–36}

Physicians also noted that these bills will contribute to greater risk for physical inactivity, children being obese or overweight, and cardiovascular disease risk, which are known health risk factors among trans youth.^{37–39} Furthermore, providers believed that these bills might increase the risk of discrimination and stigmatization of trans youth in school and on sports teams. Trans youth already report high rates of discrimination and bullying in school and on sports teams, which may be exacerbated by bills that ban their participation in gender-segregated sports in a manner consistent with their gender identity.^{17,40}

Many providers described the need to increase mental health services at their clinic to meet an increased need among their patients should these bills pass in their states. Providers who noted that they already see trans youth who participate in sports describe constraints in their ability to practice evidence-based medicine by not being able to recommend continuation in sports and other gender-segregated physical activities. This is especially relevant because of the strong scientific evidence for the mental and physical health benefits of sports participation,^{1–9,12} and providers' professional obligation to recommend health-promoting behaviors to their patients.

All providers in this study believed that these bills should not be passed. Most providers believed that it was their duty to advocate against these bills and educate lawmakers and the public about their harmful effects. Trans health advocates have often called on professional organizations to publicly oppose bills that target trans populations^{34,41} and the providers in our study also expected medical organizations along with collegiate and professional sports organizations to publicly oppose these bills, such as the Pediatric Endocrine Society and the National Collegiate Athletic Association.

Furthermore, providers believed that their profession should be actively engaged in setting participation guidelines for trans athletes that balances competition and inclusion. Providers were particularly concerned that these laws were aimed at trans youth, many of whom are taking puberty-blockers, which suppress androgen production or block its actions, and would dampen any androgen-conferred advantage in competition. Considering this fact, many providers believed that these bills were unscientific.

Limitations

Our findings should be considered in light of the following limitations. We recruited participants using email list-

serves and network referrals; thus, this is a convenience sample of providers who are well-connected to professional organizations that serve trans youth and cannot be generalized to all those who provide gender-affirming care to trans youth.

In addition, our sample was overwhelmingly composed of White cisgender women. Future research is needed to understand the perspectives of providers of color, those who are trans identified, and cisgender men. Additionally, the survey instrument states, "Several state legislatures in the United States are currently considering or have passed laws that would ban transgender and gender diverse athletes from participating in gendered sports (i.e., banning a transgender girl from a girls only team)." This language lacks precision, given that many bills do not ban participation outright. Rather, most of these bills specify that athletes competing for female-only teams must have been assigned female at birth. Given the inability to follow up on the providers' responses, we were unable to make meaningful comparisons between providers with different forms for primary training. We believe that the field would benefit from research that elicits the opinions from providers of a specific specialty and conduct in-depth interviews to better understand the variation of opinions among providers.

Despite these limitations, to our knowledge, this work is among the first to explore the perspectives of medical providers regarding legislation that would restrict the participation of transgender youth in sex-segregated sports. Results from this geographically diverse sample indicated national agreement on both the importance of sports participation for trans youth and the fear of increased discrimination, stigma, and risk for poor health outcomes among trans youth if legislation restricts access to gender-segregated sports in a manner consistent with their gender identity.

Conclusions

Overwhelmingly, medical providers of gender-affirming care for trans youth across all 50 states opposed bills that restrict their trans patients from participating in gender-segregated sports in a manner consistent with their gender identity. Specifically, providers believed that these bills would increase the risk of discrimination, suicide, and other serious mental and physical health conditions among trans youth and are therefore not founded in science. Given the numerous benefits of sports participation to the well-being of youth, providers viewed these bills as an impediment to ensuring that their patients are healthy and supported in their gender identity. Findings underscore the importance of including pediatric medical experts in policy and advocacy efforts regarding gender-segregated sports participation for trans youth.

Acknowledgments

The authors thank all the providers who participated in this study. Furthermore, they thank Drs. Shelby Davies and Jeffrey Eugene, as well as physicians from UPMC Children's Hospital of Pittsburgh for their feedback on the survey.

Authors' Contributions

L.D.H. with N.D., K.M.K., D.O., and K.E.G. conceived of the project, developed the survey, analyzed

the data, and wrote and revised this article. C.R. aided in data interpretation and writing and revising the article. All authors approved of the published version and hold themselves accountable for the accuracy and integrity of this work.

Disclaimer

The opinions expressed in this article are the authors' own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

This work was supported by the Graduate Student Research Grant Fund from the Rackham Graduate School at the University of Michigan. Dr. Kidd was supported by the National Center for Advancing Translational Science of the NIH, award number TL1TR001858 (PI Kraemer). Dr. Dowshen's work is supported by the Stoneleigh Foundation.

References

1. Logan K, Cuff S: Council on Sports Medicine and Fitness. Organized sports for children, preadolescents, and adolescents. *Pediatrics* 2019;143:e20190997.
2. Hardy LL, O'Hara BJ, Rogers K, et al.: Contribution of organized and nonorganized activity to children's motor skills and fitness. *J Sch Health* 2014;84:690–696.
3. Silva G, Andersen LB, Aires L, et al.: Associations between sports participation, levels of moderate to vigorous physical activity and cardio respiratory fitness in children and adolescents. *J Sports Sci* 2013;31:1359–1367.
4. Drenowatz C, Steiner RP, Brandstetter S, et al.: Organized sports, overweight, and physical fitness in primary school children in Germany. *J Obes* 2013;2013:935245.
5. Hebert JJ, Klakk H, Møller NC, et al.: The prospective association of organized sports participation with cardiovascular disease risk in children (the CHAMPS study-DK). *Mayo Clin Proc* 2017;92:57–65.
6. Pate RR, Trost SG, Levin S, Dowda M: Sports participation and health-related behaviors among US youth. *Arch Pediatr Adolesc* 2000;154:904–911.
7. Logan K, Lloyd RS, Schafer-Kalkhoff T, et al.: Youth sports participation and health status in early adulthood: A 12-year follow-up. *Prev Med Rep* 2020;19:101107.
8. Eime RM, Young JA, Harvey JT, et al.: A systematic review of the psychological and social benefits of participation in sport for children and adolescents: Informing development of a conceptual model of health through sport. *Int J Behav Nutr Phys Act* 2013;10:98.
9. Vella SA, Swann C, Allen MS, et al.: Bidirectional associations between sport involvement and mental health in adolescence. *Med Sci Sports Exerc* 2017;49:687–694.
10. Toomey RB, Syvertsen AK, Shramko M: Transgender adolescent suicide behavior. *Pediatrics* 2018;142:e20174218.
11. Kosciw JG, Greytak EA, Zongrone AD, et al.: *The 2017 National School Climate Survey: The Experiences of Lesbian, Gay, Bisexual, Transgender, and Queer Youth in Our Nation's Schools*. New York, NY: GLSEN, 2018.
12. Clark CM, Kosciw JG: Engaged or excluded: LGBTQ youth's participation in school sports and their relationship to psychological well-being. *Psychol Sch* 2022;1:95–114.
13. Baiocco R, Pistella J, Salvati M, et al.: Sports as a risk environment: Homophobia and bullying in a sample of gay and heterosexual men. *J Gay Lesbian Ment Health* 2018;22:385–411.
14. Freedom for All Americans: Legislative tracker: Anti-transgender student athletics. 2021. Available at <https://freedomforallamericans.org/legislative-tracker/medical-care-bans/> Accessed April 2, 2021.
15. Movement Advancement Project: Bans on transgender youth participation in sports. 2021. Available at https://www.lgbtmap.org/equality-maps/sports_participation_bans Accessed April 2, 2021.
16. An act relating to school athletics, recreation, athletic teams, and sports: S.B. 140, 32nd Leg., 1st Sess. (Alaska 2021). Available at www.akleg.gov/basis/Bill/Detail/32?Root=SB%20140 Accessed January 3, 2022.
17. Fairness in Women's Sports Act. H.B 500, 65th Leg. (Idaho 2020). Available at <https://legislature.idaho.gov/wp-content/uploads/sessioninfo/2020/legislation/H0500.pdf> Accessed March 15, 2022.
18. Save Women's Sports Act: H.B. 112, 67th Leg. (Mont. 2021). Available at <https://leg.mt.gov/bills/2021/billpdf/HB0112.pdf> Accessed January 3, 2022.
19. Relative to an exception to the opportunity for public education without discrimination: H.B. 198, 1st Year, 167th Sess. of the Gen. Court (N.H. 2021). Available at <https://legiscan.com/NH/text/HB198/id/2227419> Accessed January 3, 2022.
20. Fairness in women's sports act: S.B. 156, 2021 Reg. Sess. (La. 2021). Available at <https://legiscan.com/LA/bill/SB156/2021> Accessed January 3, 2022.
21. Relating to requiring public school students to compete in interscholastic athletic competitions based on sex and a study conducted by the University Interscholastic League on the effects of allowing a student to participate in an athletic competition designated for the sex opposite to the student's sex: S.B. 29, 87th Leg. (Tex. 2021). Available at <https://legiscan.com/TX/bill/SB29/2021> Accessed January 3, 2022.
22. International Olympic Committee. IOC framework on fairness, inclusion and non-discrimination on the basis of gender identity and sex variations. Lausanne, SU; 2021. Available at https://stillmed.olympics.com/media/Documents/News/2021/11/IOC-Framework-Fairness-Inclusion-Non-discrimination-2021.pdf?_ga=2.195521836.1048075235.1637092563-834742310.1637092563 Accessed March 15, 2022.
23. Jones BA, Arcelus J, Bouman WP, Haycraft E: Sport and transgender people: A systematic review of the literature relating to sport participation and competitive sport policies. *Sports Med* 2017;47:701–716.
24. Kavoura A, Kokkonen M: What do we know about the sporting experiences of gender and sexual minority athletes and coaches? A scoping review. *Int Rev Sport Exerc Psychol* 2020;14:1–27.
25. Westbrook L, Schilt K: Doing gender, determining gender: Transgender people, gender panics, and the maintenance of the sex/gender/sexuality system. *Gend Soc* 2014;28:32–57.

26. Coleman E, Bockting W, Botzer M, et al.: Standards of care for the health of transsexual, transgender, and gender-nonconforming people, Version 7. *Int J Transgend* 2012;13:165–232.
27. Vilain E, Betancurt JO, Bueno-Guerra N, Martinez-Patiño MJM: Transgender athletes in elite sport competitions: equity and inclusivity. In: Anderson E, Travers A (eds). *Transgender Athletes in Competitive Sport*. New York, NY: Routledge, 2017, pp. 156–170.
28. Dubon ME, Abbott K, Carl RL: Care of the transgender athlete. *Curr Sports Med Rep* 2018;17:410–418.
29. Thomas DR: A general inductive approach for analyzing qualitative evaluation data. *Am J Eval* 2006;27:237–246.
30. de Vries ALC, Steensma TD, Doreleijers TAH, Cohen-Kettenis PT: Puberty suppression in adolescents with gender identity disorder: A prospective follow-up study. *J Sex Med* 2011;8:2276–2283.
31. Turban JL, King D, Carswell JM, Keuroghlian AS: Pubertal suppression for transgender youth and risk of suicidal ideation. *Pediatrics* 2020;145:e20191725.
32. Valentine SE, Shipherd JC: A systematic review of social stress and mental health among transgender and gender non-conforming people in the United States. *Clin Psychol Rev* 2018;66:24–38.
33. Pacey MS, Dikitsas ZA, Greenwood E, et al.: The perceived health implications of policies and rhetoric targeting transgender and gender diverse youth: A community-based qualitative study. *Transgend Health* 2021. [Epub ahead of print]; DOI: 10.1089/trgh.2021.0125.
34. Kidd KM, Sequeira GM, Paglisotti T, et al.: “This could mean death for my child”: Parent perspectives on laws banning gender-affirming care for transgender adolescents. *J Adolesc Health* 2020;68:1082–1088.
35. Price-Feeney M, Green AE, Dorison SH: Impact of bathroom discrimination on mental health among transgender and nonbinary youth. *J Adolesc Health* 2021;68:1142–1147.
36. Hughes LD, Kidd KM, Gamarel KE, et al.: “These laws will be devastating”: Provider perspectives on legislation banning gender-affirming care for transgender adolescents. *J Adolesc Health* 2021;69:976–982.
37. Bishop A, Overcash F, McGuire J, Reicks M: Diet and physical activity behaviors among adolescent transgender students: School survey results. *J Adolesc Health* 2020;66:484–490.
38. Grammer AC, Byrne ME, Pearlman AT, et al.: Overweight and obesity in sexual and gender minority adolescents: A systematic review. *Obes Rev* 2019;20:1350–1366.
39. Strenn Q, Mihalopoulos NL: Cardiovascular disease risk in transgender youth. *J Adolesc Health* 2018;62:S46–S47.
40. Garofalo R, Deleon J, Osmer E, et al.: Overlooked, misunderstood and at-risk: Exploring the lives and HIV risk of ethnic minority male-to-female transgender youth. *J Adolesc Health* 2006;38:230–236.
41. Stryker S: *Transgender History: The Roots of Today's Revolution*, 2nd ed. New York, NY: Seal Press, 2017.

Address correspondence to:
 Landon D. Hughes
 Department of Health Behavior and Health Education
 University of Michigan School of Public Health
 1415 Washington Heights
 Ann Arbor, MI 48109
 USA

E-mail: landonh@umich.edu

Chapter 2

Armstrong N, McManus AM (eds): The Elite Young Athlete.
Med Sport Sci. Basel, Karger, 2011, vol 56, pp 23–46

Physiology of Elite Young Female Athletes

Alison M. McManus^a · Neil Armstrong^b

^aInstitute of Human Performance, University of Hong Kong, Hong Kong, SAR, China; ^bChildren's Health and Exercise Research Centre, University of Exeter, Exeter, UK

Abstract

The participation of girls in elite sport has increased exponentially over the past 30 years. Despite these increases a tradition for recruiting boys for exercise studies persists and our knowledge of the physiologic response to exercise in girls remains limited. Girls' physiology varies with age and maturation and is underpinned by a divergent hormonal milieu which begins early in foetal life. Sexual dimorphism underlies much of the physiologic response to exercise, and becomes most acute during adolescence when boys become taller, heavier, less fat and are more muscular than girls. Young girl athletes are not simply smaller, less muscular boys. The widening sex disparity in responses to exercise during puberty cannot always be accounted for by size. The woeful number of studies on girls and our prior inability to non-invasively study the complexity of the cellular metabolic response to exercise means an integrative understanding of girls' physiological responses to exercise remains elusive. Success in elite sport requires intense training, which for a long time was thought to cause disruption to normal growth and maturation. It would appear that exercise training, without other predisposing factors, is unlikely to cause aberrations to either growth or maturation. Nevertheless, there is clear evidence of a boundary between healthy and unhealthy levels of exertion when coupled with caloric limitation. Sports in which intense training is combined with the need for leanness may predispose girls to increased risk of skeletal and reproductive health problems, and ensuring risk is minimised should be a priority.

Copyright © 2011 S. Karger AG, Basel

The young female athlete is unique. She stands out from her peers, who show declining levels of

physical activity from early puberty [1]. In comparison to boys, she remains under-represented in competitive sport [2]. This persists to Olympic level competition with 1,704 fewer women competing at the 2008 Beijing Olympics than men [3]. Moreover, there are fewer scientific data on physiologic issues associated with exercise in girls than in boys. Traditionally, boys have been recruited for exercise studies and a search of key databases shows that this persists with comparatively fewer articles investigating physiologic responses to exercise in girls. This preference for recruiting boys reflects social constraints which can be traced back to Victorian values surrounding women, exercise and health [4]. Central to the issue of women's involvement in sport was biology, bolstered by the idea that a woman's structural and functional ability was unable to tolerate strenuous exercise, presenting considerable risk to her reproductive health. Indeed, whether or not chronic training causes less than optimal structural and functional alterations in girls remains the topic of lively debate [5–8].

Sexual dimorphism does indeed underlie much of the physiological response to acute and chronic exercise, and although a myriad of factors have been shown to influence the development of sport performance [9], structural and functional capacity represent a significant contribution to the gender differences notable. Figure 1

DEFENDANT'S
EXHIBIT

31

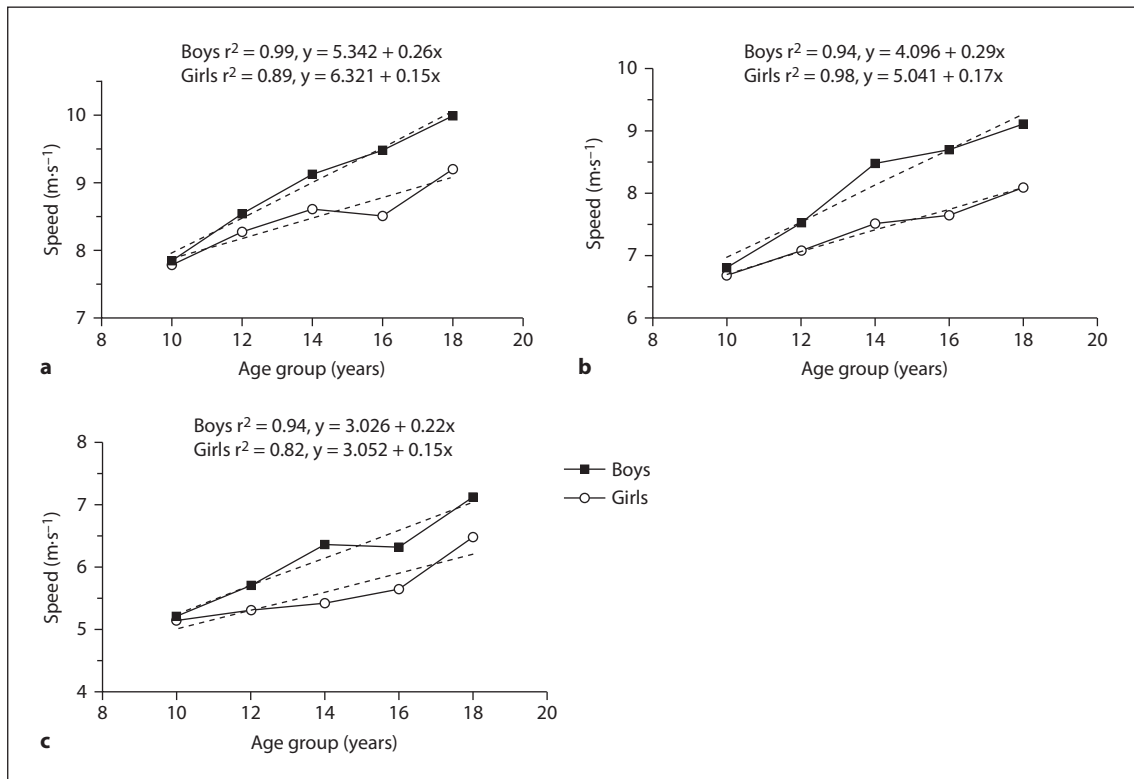


Fig. 1. Average running speed from North American age-group track and field records and world junior records over 100m (a), 400m (b) and 1,500m (c) for girls and boys.

illustrates the average running speed over 100, 400 and 1,500 m for girls and boys competing in age-group track and field championships in North America and world junior records [10, 11]. Analyses of these data show that regardless of distance, average running speed increases with age in both girls and boys ($r^2 > 0.82$ for all distances). Prior to 11 years of age differences in average speed are minimal, thereafter disparity becomes more pronounced with boys 8–15% faster than girls from 13 years of age onwards. This is consistent with the 9–15% advantage in speed adult men have over women across a wide range of distances from 100 m to 200 km [12]. When the slopes of the regression lines were compared between the sexes by age, these were significantly

different for 100 and 400 m with the rate of improvement greater in the boys than the girls (see fig. 1). The rate of improvement in running speed over 1,500 m was similar between boys and girls with a pooled slope of 0.19. That an interaction between distance, speed developmental trajectories and sex exists suggests intriguing physiological mechanisms.

A key question to be addressed is whether the superior performances of boys relate to qualitative discrepancies in functional capacity or alternatively, they are simply a function of the increased body size and disparate body composition that accompanies adolescent growth and maturation. Appropriate adjustment for differences in body size and composition dampen many apparent

physiological differences [13, 14]; however, young girl athletes are not simply smaller, less muscular boys. Girls' physiology varies with age and is underpinned by a divergent hormonal milieu which begins early in foetal life. There is evidence that in the two-cell stage of embryonic development, long before visual gonadal differentiation, the sex-determining region of the Y chromosome has already been transcribed [15]. Testosterone secretion commences at about 3 months in the male foetus, the absence of which in the female foetus allows maturation of the female reproductive organs, and by birth subtle differences in cardiac function and body composition already exist [16, 17]. The accentuated hormonal adjustments which occur during adolescence result in differential development and a widening sex disparity in many physiological responses to exercise, some of which cannot be accounted for by size.

The primary focus of this chapter is on physiologic issues that are associated with exercise in girls; but to illustrate sexual dimorphism, comparisons will be made with boys where relevant. The chapter begins with a brief overview of growth and maturation in girls, including a discussion of issues related to body composition and size. This is followed by a focus on the acute responses to sustained aerobic exercise, as well as short-duration high-intensity exercise in girls. The possible physiological mechanisms that underlie these responses, such as sex differences in pulmonary, cardiac, and peripheral function, as well as cellular metabolism are discussed. The chapter concludes with consideration of the contention that intensive training poses a substantial threat to the development and health of young girls.

Growth and Maturation

Growth hormone (GH), insulin-like growth factor I (IGF-I), the sex steroids and insulin are all potent anabolic hormones. Their complex interactions enable linear growth, bone mineralization, increases

in muscle and metabolic adaptations during childhood and adolescence and much of this hormonal mélange is sex dependent [18]. For example, the physiological effects of the sex steroids testosterone and oestrogen differ markedly, with evidence that combined testosterone and GH administration causes increases in IGF-I concentrations, resulting in enhanced anabolism, greater increases in fat free mass and higher whole body protein synthesis in boys [19]. Oestrogen administration, on the other hand, has been shown to have no effect on whole body protein synthesis in girls [19]. Sex dimorphic growth and development is most pronounced during adolescence, which forms the primary focus of the following sections.

Stature and body mass follow a double-sigmoid growth pattern in girls and boys, with rapid gains in infancy, slower yearly gains of about 5–6 cm in stature and 2.25–2.75 kg in body mass through childhood, and a second rapid gain in adolescence [20]. Girls usually begin adolescent growth before boys and progress at a faster rate than boys [21, 22]. At the peak of the adolescent growth spurt, girls gain approximately 8–9 cm in a year in stature. Boys only gain about 3 cm more in stature during the adolescent growth spurt, but are about 11–13 cm taller by adulthood because of their extra pre-adolescent growth [23]. From onset to completion, adolescent growth in stature lasts about 4–4.5 years in girls, until rising levels of oestrogen induce epiphyseal fusion marking an end to the growth in stature, usually around a skeletal age of 15 years [18, 21]. Peak body mass velocity is in lag by some 4–6 months with peak height velocity and total body mass gains of 16 kg are usual during the adolescent growth spurt in girls [24].

Other body proportions such as sitting height, leg length, biacromial and bicristal breadths follow a similar growth pattern to stature. Leg length and sitting height differ little between girls and boys during childhood [23]. At the onset of adolescence rapid growth in leg length precedes trunk growth. Boys surpass girls in leg length by about 12 years of age and in sitting height around

14 years of age. Yet, the ratio of sitting height to stature is higher in girls than boys through adolescence indicating relatively shorter legs in girls for the same stature. Girls have a marginally wider bicristal breadth than boys from late childhood to late adolescence, when boys catch-up [23]. In contrast, boys experience much more dramatic increases in biacromial breadth compared to girls [23]. When bicristal breadth is expressed as a ratio of biacromial breadth (hip-to-shoulder ratio), in comparison to boys values are higher in girls from early childhood, with bicristal breadth approximately 72–73% of biacromial breadth, remaining quite stable through adolescence. In boys a decline in this ratio is noted from about 70% at 11 years of age, to 65% by 16 years of age, which is an outcome of the disproportionately faster growth of biacromial breadth [23]. Interestingly, although stature as well as mass-to-stature ratio differ between girls involved in different competitive sports, minimal differences in other proportions such as arm span and seated height have been reported [25, 26]. Greater mass-to-stature ratios can confer performance benefit in some sports such as throwing events; however, the combined effect of broader hips and shorter legs that usually accompany a greater mass-for-stature and characterize early maturation in girls, is generally disadvantageous. Although data are sparse, young female athletes in running events or gymnastics are generally more likely to be characterized by longer legs, lower hip-to-shoulder ratios and lower mass-for-stature.

Assessment of maturity stage is vital but poses considerable challenge. Skeletal age is the biological marker of choice, but is hindered by ethical constraints related to ionizing radiation exposure. The timing and tempo of sexual maturation in girls has most commonly been described using the visual descriptive stages of secondary sexual characteristics. These were first documented by Reynolds and Wines [27], and then further refined by J.M. Tanner and are better known as Tanner stages [28]. There is a large normal variation in the timing and tempo of sexual maturation in girls, as

well as clearly documented sex differences. Like linear growth, girls normally begin sexual maturation before boys and progress toward full maturity at a faster tempo than boys [22].

A recent large-scale longitudinal study of Caucasian and African-American children suggests that the average girl begins breast development at 9.8 years, whilst the average boy begins genital development at about 10.3 years [21]. Pubic hair growth usually occurs around 10.2 years in girls and around 11.3 years in boys. The onset of the initial stages of sexual maturity in these American girls is somewhat earlier than previously published data for European girls, for whom breast budding was reported to occur at about 10.5 years and Tanner Stage two for pubic hair at 10.8 years [28, 29]. This discrepancy probably reflects the different racial mix of the groups, in addition to a possible secular trend for a declining age of onset of maturation [30].

Asynchronous maturation of secondary sex characteristics in girls is common and has been defined as a difference of at least 4 months between breast and pubic hair development. About 51–66% of girls follow an asynchronous maturation pattern [21, 31]. Most (about 70%) follow a thelarchal pathway, with breast development beginning prior to pubic hair growth (adrenarche). Asynchrony usually persists into Tanner stage 3, the onset of which is on average 11.3 years for breast development, with pubic hair stage three occurring some 2 months later [31]. As the latter stages of sexual maturity are attained, development becomes more synchronous. A minority of girls (approximately 30%) follow an adrenarchal asynchronous pathway, in which pubic hair development precedes breast development. Thelarchal asynchrony is believed to result from initial advanced stimulation of gonadotropin and oestrogen, thereby enabling earlier breast development. The converse is true of adrenarchal asynchrony, where the advanced production of testosterone and adrenal hormones promotes earlier pubic hair growth. Age of onset of menses usually

occurs during Tanner stage three for breast development. For those girls following the thelarchal pathway, menses occurs at an earlier age, around 12.6 years. Those following the adrenarchal pathway usually begin menses around 13.1 years [21, 31]. Lower oestrogen levels are noted in girls following an adrenarchal maturation pathway which persists throughout adolescence. This affords a body composition advantage, characterized by a lower sum of skinfolds, percent body fat and waist-to-hip ratio [31]. Girls involved in intensive training are generally characterized by lower percent body fat, but there is little evidence to suggest that they preferentially follow an asynchronous adrenarchal pathway [32].

The presence of asynchronous sexual maturation has implications when comparing young athletes with non-athletes, as well as making comparisons between girls and boys. Whether alignment is on the basis of a single marker (e.g. pubic hair development), the creation of a composite score for both pubic hair and breast development, or on differing secondary sex characteristics (such as genitalia and breast development), the assumption is that the timing of the appearance of a particular characteristic, as well as the tempo, is homogeneous. This is clearly not always the case, and it has been suggested that alignment of sexual maturation to other biological or somatic markers of maturation (e.g. age at menarche or peak height velocity) is more appropriate [33]. Menarchal age is convenient if retrospective, otherwise, like peak height velocity, a prospective research design is necessary. Assessment of maturational stage continues to present a real methodological challenge to paediatric exercise physiologists.

Body Composition and Size

Fat

Small sex differences in fat mass and percent body fat are evident from mid-childhood, with levels

rising substantially in girls during adolescence. Body fat gains by the end of puberty usually result in 26–31% body fat in the average adolescent girl [34, 35]. Young athletes are generally leaner than the average non-athletic girl, but this is dependent on the chosen sport. Values as low as 14.3% have been reported for 15-year-old rhythmic gymnasts, with gymnasts generally showing lower body fat than other athletes [34]. Body fat values from 21 to 25% have been reported for dancers, distance runners and cross-country skiers from the ages of 10–17 years [36–39]. Higher values have been reported for 13- to 17-year-old high-school athletes competing in lacrosse, soccer, softball, swimming, track and field and volleyball (mean $27.4 \pm 0.7\%$) [40].

Sex steroids are major determinants of body fat distribution, with the increases in body fat generally subcutaneous and in the gluteal and femoral regions in girls. Fat mass, combined with a smaller leg length-to-stature ratio, lowers the centre of gravity in girls, thereby affording better balance. However, fat mass is also negatively related to heat dissipation, which may prove disadvantageous in girls during endurance events in hot environmental conditions [41].

Fat tissue has relatively uniform properties throughout life, with negligible water content and a tissue density of $0.9007 \text{ kg} \cdot \text{l}^{-1}$ [42]. Recent reference data from Wells et al. [42] have shown that in comparison lean tissue shows sex-specific chemical maturation, with decreases in water content and increases in density with increasing age. These new data have implications for the assessment of body fat since previous reference data extrapolated rather than directly assessed age specific tissue densities and hydration. Wells et al.'s [42] work provides the first comprehensive empirical data set for lean tissue properties for 4- to 23-year-old boys and girls, with lean tissue density values in girls of $1.0905 \text{ kg} \cdot \text{l}^{-1}$ at 8–9 years, rising to $1.1021 \text{ kg} \cdot \text{l}^{-1}$ at 16–17 years. Lean tissue hydration values declined with age from 75.2% at 8–9 years to 73.7% at 16–17 years

in girls. Importantly, this study has shown that these new values differ from previously simulated values. The lean tissue density values of Wells et al. [42] were consistently higher, whilst the hydration data were consistently lower than those reported by Lohman [43]. Comparisons of % fat calculated from densitometry with the Lohman [43] formula led to a between-study error of -1 to 2.5% fat in the average girl. Wells et al. [42] provide important new reference values for the assessment of body fat by both hydrometry (total body water, bioelectrical impedance) and densitometry which should ensure greater clarity in future analyses.

Muscle

At birth, boys tend to have a greater lean mass than girls. This difference remains small but detectable throughout childhood with about a 10% greater lean mass in boys than girls prior to puberty [17]. The sharp increase in muscle mass disparity between the sexes during puberty indicates a primary role of the gonadal steroidal hormones. Muscle mass in girls increases from about 25 kg at 10 years of age to about 45 kg by 18 years of age [42]. Reported values for 15- to 17-year-old female athletes are not dissimilar, ranging from 42 to 53 kg [23]. These gains in muscle tissue represent an increase of about 5% in muscle mass. The relative contribution of muscle mass to total body mass usually declines once consideration is given to the relative contribution of fat mass. In comparison, the androgen-mediated growth of muscle in boys results in muscle mass reaching about 55% of total body mass at maturity [44]. The greater overall skeletal muscle mass in adolescent boys creates a potential cascade of functional differences apparent in adults such as differing muscle fibre size, activities of metabolic enzymes, lipid content and oxidation, relative expression of myosin isoforms, and fatigability [45–50]. Maturation of these features remains poorly understood.

The use of ultrasonography and magnetic resonance imaging (MRI) are providing insight into changes in muscle architecture with growth. A recent study using ultrasonography demonstrated that muscle thickness (a marker of physiological cross-sectional area) and pennation angle were correlated with age from 4 to 10 years in both sexes [51]. Findings from MRI studies have shown similarly that muscle cross-sectional area increases with age from childhood through adolescence, and more so in boys than girls [52]. Pennation angle on the other hand has not been found to differ between the sexes [51]. Whilst muscle cross-sectional area and pennation angle are related to age, this has not been shown in muscle fibre length [51]. Muscle fibre length has been found to have high inter- and intra-individual variation, which may reflect a greater malleability in response to external stimuli such as the extent and intensity of exercise [51].

Morphological change in the muscle impacts upon function. Maximal strength, for example, is dependent on the specific joint angle (force-length relationship), contraction type, muscle cross-sectional area and velocity. The length of the muscle fibre is proportional to the absolute maximum contraction velocity, whilst the pennation angle dictates the proportion of force transmitted to the tendon. Muscle strength, expressed as torque, increases with age in children, but gains are greater in boys. This has been presumed to be an outcome of the greater muscle cross-sectional area [53]. Alternatively, there may be intrinsic sex differences in the fibre composition and fatigue characteristics of skeletal muscle that materialize during adolescence that also influence the ability to increase torque.

In adults, several studies have reported higher glycolytic enzyme activity and lower oxidative enzyme activity in men compared to women, supporting the contention that men have a lower proportion of type I fibres [50, 54]. Data on muscle fibre typing in children are limited because of the invasive nature of the biopsy methodology,

but there is evidence to show that differentiation of fibre type occurs during the first few years of life. About 10% of skeletal muscle fibres remain undifferentiated up until puberty, with no sex difference notable in the percentage of type I fibres (slow-twitch oxidative fibres) during childhood [55]. By adolescence females have a lower % of type I muscle fibres than males [45, 56] and the type II muscle fibres of young men are bigger than their type I fibres, something not evident in young women [56, 58].

Although boys gain more in strength than girls during adolescence, elite girl athletes are stronger than their less athletic peers. For example, average quadriceps and biceps isometric strength was reported to be 22% greater in elite gymnasts and swimmers and 18% greater in tennis players compared to less athletic school children [59]. Interestingly this study found no differences in strength between sports in the girls, even when co-varied for body mass. The relationship between strength and body mass, or strength-to-mass ratio, has been seen as an important predictor of sport performance particularly in gymnastics, middle- and long-distance running. Indeed, elite adult women runners such as Yvonne Murray and Greta Waitz were 17–18% below the average body mass for their stature at the peak of their running careers, which suggests relative strength was high. The work of Bencke et al. [60] has shown that 11-year-old girl gymnasts were the smallest, lightest and possessed the highest explosive strength compared to other athletes, suggesting high relative strength confers advantage in some sports in girls.

Bone

Bone characteristics differ little between boys and girls prior to puberty, but then follow two sex-divergent growth paths. During the adolescent growth spurt boys experience increases predominantly in bone diameter and cortical thickness

due to periosteal apposition [17]. Girls on the other hand experience increases in cortical thickness, a decrease in medullary diameter, and little increase in periosteal diameter as a result of oestrogen inhibition of periosteal apposition [17, 61]. It should be noted that bone accretion and endocortical features appear to be site specific with data showing endocortical resorption at the mid-femur and proximal tibia in girls through puberty, but no endocortical resorption at the radial diaphysis [62, 63].

During puberty, bone mineral content (BMC) accrual rate is in lag with muscle accrual rate, suggesting that muscle enlargement, and concomitant increases in muscle force, are important for bone development [64]. Indeed, the ‘functional muscle-bone unit’ hypothesis suggests muscle force is a primary determinant of bone mass, structure and strength [65]. Young female runners and gymnasts have been shown to have elevated bone mass and enlarged bone size at specific sites such as the radius and lumbar spine in gymnasts [66] and the femur in runners [67], reflecting the specific mechanical-loading patterns these sports require. This has led some to conclude that muscular force alone explains the impact loading effect on bone [68–70]. On the contrary, recent research has shown that bone mass, size and strength increases in the upper extremity in gymnasts are independent of maturation, stature and muscle cross-sectional area and substantiates the hypothesis that other non-muscular loading factors may also account for skeletal adaptations [71, 72].

Puberty is the most favourable period for augmented bone mineralization, with about one quarter of adult bone being laid down. Bone mineral accrual is sex and maturity dependent and appears to be enhanced by oestrogen. It is clear that the early pubertal and pre-menarchal years are particularly important for young girls in terms of optimizing their bone mineralization and weight-bearing exercise plays a key complementary role in this process [73].

Body Size Considerations

By the time the adolescent growth spurt is complete the body size, shape and composition of boys and girls is different. Boys have become taller, have longer legs, broader shoulders, are heavier, and have less fat and more muscle than girls. The effect of these discrepancies on performance is substantial, and it is important in understanding girls' physiologic responses to exercise that we are able to effectively partition the impact of size from function. Traditionally in exercise physiology this has been achieved by expressing the physiological measure of interest (y) as a ratio of an appropriate marker of body size (x) to give the ratio y/x . Tanner suggested in 1949 that the use of such ratio standards to scale physiological measurements to size was 'theoretically fallacious and unless in exceptional circumstances, misleading' [74]. Yet this has largely been ignored with much of the comparison between men and women, or boys and girls based on ratio standards [75, 76]. An implicit assumption with the ratio standard is that the relationship is linear and the y intercept is zero. Additionally, ratio standards should only be used when the coefficient of variation (V) for body size (x), divided by the coefficient of variation (V) for the physiological variable (y), equals the Pearson product moment correlation coefficient (r) for the two variables, expressed by the equation $V_x/V_y = rx/y$. These assumptions are rarely met and the outcome is scaling distortion, which may have obscured our understanding of the physiologic responses of girls [77].

Theoretically, morphological and physiological variables are scaled according to the general allometric equation $y = ax^b$, where y is the morphological or physiological variable of interest, x is the chosen size denominator, b is the scaling exponent and a is the constant [78]. When this equation is solved the resultant power function ratio (y/x^b) is derived. Various studies have shown that with careful consideration of the denominator, alternative approaches, such as the allometric

power function ratio or more complex multilevel modelling of longitudinal data, are more appropriate than ratio scaling when comparisons of various physiological outcomes between individuals of differing body size are sought [79, 80]. These alternatives should, wherever possible, be utilised.

Acute Responses to Aerobic Exercise

Peak oxygen uptake (peak $\dot{V}O_2$), the highest $\dot{V}O_2$ elicited during an exercise test to exhaustion in children, is well-established as the best single measure of aerobic fitness [81]. In comparison to boys, girls are characterised with a smaller absolute peak $\dot{V}O_2$. Predicted values range from 1.5 to 2.2 litres $\cdot$ min⁻¹ in 10- to 16-year-old girls and are lower than boys by 11, 19, 23 and 27% at ages 10, 12, 14 and 16 years of age, respectively [82]. Peak $\dot{V}O_2$ is strongly correlated with body size and composition and thus, much of the divergence in values reflects this. When expressed as a ratio standard with body mass (ml $\cdot$ kg⁻¹ $\cdot$ min⁻¹), peak $\dot{V}O_2$ shows a progressive decline in girls from 13 years of age, with values dropping from approximately 45 to 35 ml $\cdot$ kg⁻¹ $\cdot$ min⁻¹ [83]. In contrast, mass-related peak $\dot{V}O_2$ in young female runners has been found to be relatively constant with values of 56.3, 57.1, 56.9 and 54.3 ml $\cdot$ kg⁻¹ $\cdot$ min⁻¹ at ages 10, 12, 14 and 16 years, respectively [84]. Likewise, peak $\dot{V}O_2$ has been shown to be fairly stable between 11 and 16 years of age in elite girl swimmers and tennis players [85], with values of 51–52 ml $\cdot$ kg⁻¹ $\cdot$ min⁻¹ and 47–49 ml $\cdot$ kg⁻¹ $\cdot$ min⁻¹, respectively. When multilevel modelling was used to account for mass, stature and biological age, the elite girl swimmers and tennis players showed increases in peak $\dot{V}O_2$ until late puberty when increases became non-significant [85]. Similarly, in the less athletic population when more appropriate allometric adjustment is used to partition size effects in body mass and stature, peak $\dot{V}O_2$ has been found to increase significantly from 11 to 13

years in girls, and then remain constant with no decline into adulthood evident [81].

Dramatic pubertal changes in muscle, fat and mass contribute to the widening of the sex difference in peak $\dot{V}O_2$. When a marker of body fat was included in a multilevel regression model which incorporated body mass, stature and age, the sex difference in peak $\dot{V}O_2$ was reduced, but the greater increase in boys' peak $\dot{V}O_2$ with growth compared to girls was still not fully explained [83]. Equally, longitudinal data have shown that even when differences in body mass and fat mass are controlled for allometrically, girls utilise less oxygen than boys during submaximal exercise, and this becomes more pronounced with age [86]. Understanding the physiologic mechanisms that underlie these size-independent sex differences in peak and submaximal $\dot{V}O_2$ requires consideration of the coordinated systems response, which includes pulmonary, cardiac and peripheral adjustments to the demands in muscular energy. A discussion of key features of each follows.

Pulmonary

It was generally assumed that because exercise training exerts little influence on lung structure or function that the lungs exert minimal influence on oxygen transport. However, there is evidence that lung function adaptation does occur as a consequence of exercise training in girls [87]. Moreover recent investigation of sex differences in pulmonary structure and function in adults has shown considerable effects on gas exchange and the integrated ventilatory response during exercise, in particular exercise-induced arterial hypoxia [88]. There are well-documented sex differences in anatomical aspects of the pulmonary system which occur during lung growth [89]. The consequence of sex dependent pubertal thoracic growth is a larger thoracic width in boys. When coupled with a greater muscle mass for generating lower lung function, boys have approximately 25% greater lung volumes than girls

who are matched for stature [89]. By adult life, in addition to the smaller lung volumes, stature and age independent lower resting diffusion capacity (corrected for haemoglobin), lower maximal expiratory flow rates [90], and a greater occurrence of exercise-induced hypoxia has been shown in women [91]. Equally, there is also evidence that when matched for size and aerobic power women do not have reduced diffusion capacity or impaired ventilation perfusion during exercise [92].

In children, like adults, exercise pulmonary gas exchange depends on pulmonary ventilation ($\dot{V}_E$) and at maximal work rates high rates of ventilation are usual. Maximal values of 49–95 litres $\bullet$ min⁻¹ have been recorded for girls between the ages of 9 and 16 years [93] and there is a consistent sex difference with values somewhat higher in boys (58–105 litres $\bullet$ min⁻¹) for the same age span. It should be noted that cross-study comparisons are difficult given the dependence of ventilation on the protocol and data such as these need to be interpreted cautiously. Maximum ventilation remains higher in boys, whether controlled for body size using a ratio standard or allometric adjustment with either stature and/or body mass [94, 95]. Thus, the higher peak $\dot{V}O_2$ in boys is indeed supported by a higher $\dot{V}_E$.

During exercise, an expiratory flow limitation is apparent in adult women but not men, resulting in a greater oxygen cost of breathing and the onset of arterial desaturation [88, 96]. Recent evidence has provided a comparison between pre-pubertal boys and girls and found no difference in the occurrence or severity of expiratory flow limitation between girls and boys and no changes in arterial saturation during exercise to maximum [97]. Others have found little evidence of exercise induced arterial hypoxaemia in pre-pubertal girls, or lower ventilatory efficiency at maximum [98, 99]. When Armstrong et al. [94] compared ventilatory parameters during submaximal exercise at the same absolute intensities they noted that girls demonstrated higher ventilatory equivalents for oxygen and carbon dioxide, i.e. poorer ventilatory

efficiency in comparison to boys. However, when they compared submaximal ventilatory efficiency during the same relative exercise intensities, values were remarkably similar between the sexes. This suggests that differences apparent at absolute submaximal exercise intensities simply reflect the higher relative percentage of maximum that girls are working at and do not denote true inefficiency.

There is little evidence that prior to puberty pulmonary structure or function limits oxygen uptake, however, considerable evidence has shown pulmonary function influences gas exchange in adult women, suggesting that maturational adjustments occur. At present however, there is little evidence to substantiate this.

Blood Volume and Haemoglobin

Assessment of blood volume in children and adolescents is complex and the variability in techniques means there are considerable discrepancies between studies. There are conflicting results regarding changes in blood volume with age. Some have shown that blood volume per unit body mass increases with age [100], others have found no change [101], whilst others report decreasing blood volume with age [102]. Likewise, data on sex differences in blood volume between girls and boys are mixed. When normalised using a ratio standard with body mass, differences between girls and boys were apparent from about 6 years of age, with values lower in the girls [103]. In contrast, when normalised using a ratio standard for lean body mass, sex differences are no longer apparent for pre-pubertal children, or at any maturational stage [103].

In boys, haemoglobin rises through adolescence to about $152 \text{ g} \cdot \text{l}^{-1}$ by 16 years of age [104]. Girls, on the other hand, usually demonstrate a plateau in haemoglobin concentration with values of about $137 \text{ g} \cdot \text{l}^{-1}$ by 16 years of age [104]. Highly trained adolescent female athletes also

show lower haemoglobin concentration values compared to trained boys, with about a 7% difference [105]. Fully saturated, 1 g of haemoglobin carries 1.34 ml of oxygen, and one would presume that the smaller increase in haemoglobin in girls would result in a reduced oxygen carrying capacity in comparison to boys. However, it has been shown that haemoglobin concentration is not a significant predictor of peak $\dot{V}\text{O}_2$ in 11- to 17-year-olds once body size and composition and maturation have been controlled for [83].

Cardiac and Vascular Considerations

There are clear differences in cardiac function at rest and during exercise between girls and boys, with differences apparent even prior to puberty. The electrical conduction system is influenced by sex steroid hormones, with girls normally having higher resting heart rates than boys – somewhere in the magnitude of 90 beats per minute at around 10–12 years of age [106]. This is thought to relate to intrinsic differences in the sinus node pacemaker [107], a difference notable at birth with newborn boys displaying lower baseline heart rates than girls [16]. The higher resting heart rate in girls is often explained as an artefact of differences in cardiac dimensions, and indeed the ratio of heart mass to body mass has been found to be higher in boys than girls at birth, remaining so through adolescence [106]. Heart volume has also been found to be greater in boys with values of 342 and 403 ml for pre-pubertal girls and boys, respectively, and of 466 and 561 ml for pubertal girls and boys, respectively [108]. When adjusted for body mass these differences were found to persist through puberty (female $10.0 \text{ ml} \cdot \text{kg}^{-1}$; male $10.8 \text{ ml} \cdot \text{kg}^{-1}$). Inconsistencies in the findings, however, are present and others have found no differences in either left ventricular mass [109] or heart volume [110].

Echocardiographic studies that have shown greater left ventricular mass in boys compared

Table 1. Oxygen uptake, stroke index, cardiac index and arteriovenous oxygen difference at maximal cycle ergometer exercise

Reference	Sex	n	Age (years)	SI (ml·m ⁻²)	HR (bpm)	CI (litres·min ⁻¹ ·m ⁻²)	a-v O ₂ difference (ml·100 ml ⁻¹)	Peak $\dot{V}O_2$ (litres·min ⁻¹ ·kg ⁻¹)	Peak $\dot{V}O_2$ (litres·min ⁻¹)
Cumming [111]	F	29	11.8±3.1	46±3 [†]	174±11	8.61±8.1 [†]	–	–	–
	M	31	12.6±3.5	56±13	170±17	10.1±1.8	–	–	–
Rowland et al. [112]	F	24	11.7±0.5	55±9 [†]	198±9	10.9±1.7 [†]	12.3±1.9	40.4±5.8 [†]	1.84±31
	M	25	12.0±0.4	62±9	199±11	12.3±2.2	12.2±1.7	47.1±6.1	1.98±28
Obert et al. [113] Pre-training experimental group	F	7	10.66±0.3	47±7 [†]	204±5	9.4±1.4 [†]	13.2±1.6	40.9±8.9 [†]	–
	M	9	10.66±0.5	52±8	199±9	10.5±1.8	13.0±2.1	44.1±6.1	–
Pre-training control group	F	10	10.41±0.3	46±6 [†]	202±7	9.4±1.2 [†]	13.1±2.8 [†]	42.4±5.6 [†]	–
	M	9	10.5±0.3	49±5	202±7	9.7±0.8	15.6±1.5	51.5±6.3	–
Winsley et al. [114]	F	9	10.2±0.3	45±6	192±11	8.7±1.1	12.6±1.6 [†]	–	1.23±0.8 [†]
	M	9	10.1±0.5	47±8	195±11	8.9±1.4	14.8±2.1	–	1.41±1.8

SI = Stroke index; HR = heart rate; CI = cardiac index; a-v O₂ difference = arterio-venous oxygen difference; $\dot{V}O_2$ = oxygen uptake. [†] Significant differences noted.

to girls have suggested that the reduced cardiac mass in girls may be associated with reduced contractility, reduced pre-load or increased afterload [106]. All of these could result in a reduced stroke index (SI) and therefore reduced cardiac index (CI). Cardiac index has generally been found to be higher in boys than girls at maximal exercise (table 1) and in the absence of sex differences in maximal heart rate, it would appear that SI most likely accounts for this difference. Absolute maximal SI index has been reported to be between 7 and 13% less in girls than boys. When corrected for body fat, this difference was reduced to 5.2% [112], but remained nonetheless. Interestingly, the lower maximal SI index apparent in girls has not always been found to relate to left ventricular dimensions, which suggests sex differences may instead relate

to other factors such as the peripheral pump, systemic vascular resistance or differing adrenergic responses [113, 115].

Evidence of cardiac re-modelling following training has provided some insight into the role systemic vascular resistance may play in SI differences between boys and girls [113]. Following 13 weeks of training, both pre-pubertal boys and girls increased LV end-diastolic diameter and left ventricular mass. However, only LV end-diastolic diameter was related to percent increase in SI. Percent increase in SI was also inversely related to systemic vascular resistance, suggestive of vascular adaptations in response to high-intensity training. Of note, the decrease in systemic vascular resistance was greater in the boys than the girls, which may account for the greater increase in maximal SI in the boys.

The vasoregulatory capacity of the arterial and arteriolar vessels manipulates peripheral resistance. When blood is effectively distributed to the working muscle, peripheral resistance is reduced, which unloads the heart improving the capacity of the heart to increase SI. This is achieved by improving the flow of blood to and through the muscle and both the vasculature and skeletal muscle pump are involved. Interestingly, no sex differences in arterial compliance have been noted in pre- and early-pubertal children [116], although the beneficial role of oestrogen in vasodilation is well established and female advantage in arterial compliance is apparent in adults [117].

The skeletal muscle pump utilises the rhythmic muscle contractions to empty the venous vessels, aiding blood muscle hyperaemia and venous return. There is scant information on the skeletal muscle pump in children, but evidence in boys suggests, like adults, the skeletal muscle pump is associated with improved CI [118, 119]. In Rowland et al's [119] study, arterio-venous oxygen (a-v O_2) difference, a composite index of the haematological components of oxygen delivery, remained constant during unloaded exercise suggesting the increases in muscle oxygen supply were met by the increasing blood volume. Conversely, as exercise intensity increased with loading, a-v O_2 difference increased indicating decreased effectiveness of the muscle pump in satisfying the metabolic demands of the working muscle.

Whilst some studies have found no differences in estimated a-v O_2 difference at maximal or sub-maximal intensities between pre-pubertal girls and boys [112, 120], there are conflicting findings. Data recently published from a thoracic impedance measure of peak CI and MRI markers of cardiac size [114] demonstrated that pre-pubertal boys had a 16.7% higher a-v O_2 difference than girls. This was the only distinguishing factor to explain the significantly higher peak $\dot{V}O_2$ in the boys compared to the girls and unlike other studies no difference in either CI or SI were apparent

at maximal exercise. It is interesting to note that a-v O_2 difference was 16% lower in the girls of the control group (table 1) in the study of Obert et al. [113]. These findings are intriguing, but confirmatory studies are needed to help understand the inconsistencies in the extant data.

Muscle Cellular Metabolism during Moderate Intensity Exercise

Characterizing muscle metabolism during exercise is extremely challenging and for a long time hampered by the need for invasive measurement of enzymatic activity. ^{31}P magnetic resonance spectroscopy (MRS) has enabled the study of high energy phosphates non-invasively in human skeletal muscle. This technique can provide an estimation of skeletal muscle metabolic activity via examination of creatine phosphate (PCr), inorganic phosphate (P_i) and intracellular pH, and has been validated in both adults and children [121, 122]. There remain methodological challenges in the paediatric population, which have been outlined by Armstrong and Fawcner [123], but the data available are providing fascinating insight into cellular metabolic processes.

Children, like adults show high correspondence between MRS determined muscle phosphocreatine (PCr) activity and the pulmonary oxygen uptake ($p\dot{V}O_2$) kinetic response [124, 125]. This implies that $p\dot{V}O_2$ kinetics also provide a marker of energy utilization at the muscular level, one which may prove very useful in understanding the interplay between cardiopulmonary and metabolic processes during exercise. More comprehensive descriptions of oxygen uptake kinetic assessments have been provided elsewhere [126] and only the salient issues related to girls' responses are summarized here. The $p\dot{V}O_2$ kinetic response is tri-phasic, but only phases II and III pertain to muscle oxygen uptake kinetics. During moderate intensity exercise, the phase II $p\dot{V}O_2$ kinetic response involves

an exponential increase in oxygen uptake toward steady state, which signifies increases in muscle $\dot{V}O_2$. The primary response is described by a time constant (τ), representing the time taken (s) to achieve 63% of the change in $p\dot{V}O_2$. The attainment of a steady state denotes phase III. At higher workloads, i.e. those above the maximal lactate steady state, the $p\dot{V}O_2$ kinetic response alters, with phase III showing a delayed increase, eventually resulting in a $p\dot{V}O_2$ value higher than predicted on the basis of exercise intensity. This 'slow component' represents an increasing inefficiency in energy turnover and negatively correlates with increases in $\dot{V}O_2$ per unit increases in work, suggesting fatigue. To ensure confidence in the kinetic parameters estimated, the level of measurement rigour needed is high [126]. Few of the available oxygen uptake kinetics studies with children provide this and as such information on girls is very limited.

There is little evidence of a sex difference in $p\dot{V}O_2$ kinetic responses during moderate intensity exercise in children [127]. Neither have sex differences been found in boys and girls for MRS determined pH , P_i to PCr ratio (P_i/PCr) or PCr kinetic time constant at either the onset or offset of moderate intensity exercise [128]. In contrast, a study of the kinetic responses to high-intensity exercise found sex differences [129]. Results showed phase II $p\dot{V}O_2$ kinetics were approximately 20% slower in pre-pubertal girls compared to boys and the relative contribution of the $p\dot{V}O_2$ slow component to the end exercise $p\dot{V}O_2$ in the girls was about 30% greater. This is suggestive of a lower tolerance of fatigue in the girls, but the mechanisms underlying this response are not yet understood. One hypothesis suggests that these differences reflect a difference between boys and girls in the energetic profiles of the recruited muscles.

Barker et al. [130] have explored high-intensity exercise responses of the quadriceps muscle using MRS in children, but showed in accord with the $p\dot{V}O_2$ kinetic work, that girls responded with a greater anaerobic metabolic contribution than

boys. These findings were partly attributed to the inequalities in maturity status, with relatively immature boys compared to the girls. Maturation of the cellular anaerobic response was noted in the girls in this study, who progressed from a response that was attenuated prior to puberty, but adult-like with ensuing maturation. This was not apparent in the boys, most likely an artefact of the narrow age range of the boys (9–12 years). Generally, high-intensity work requires the recruitment of fast twitch muscle fibres that are faster and larger, with a greater glycolytic and lower oxidative capacity. As discussed earlier, there is evidence of sex differences in muscle fibre type and size which vary with age and maturation, and clearly comparison of cellular metabolism during high-intensity exercise in girls and boys who are more closely aligned in terms of maturation is something which deserves further enquiry.

To summarise, there are differences between boys and girls in the aerobic responses to exercise which cannot be accounted for solely by size. Ventilatory parameters do not appear to influence peak $\dot{V}O_2$ in pre-pubertal children, however, there is scant information on the maturation of ventilatory responses in girls. It has been suggested that peripheral factors may be more important in defining aerobic fitness than cardiac function [131], but these are poorly understood in children and in particular in girls.

Acute Responses to High-Intensity Exercise

Most sports require short-duration bursts of high-intensity effort, which are supported by high muscle energy turnover. The direct examination of muscular energetics during short-duration high-intensity exercise is complex and instead investigations have largely concentrated on mechanical output markers of short duration exercise performance. The most commonly employed tests are the Wingate cycle ergometer test (WAnT) and cycle ergometer force-velocity tests,

both eliciting markers of leg power. Wingate test values for leg peak power in girls aged 11–16 years have ranged from 260 to 542 W [132–136], whilst comparable values between 250 and 555 W have been recorded using force-velocity tests in similarly aged girls [137–140]. Mean power values from the Wingate test have ranged from 228 to 341 W in 11- to 16-year-old girls [132, 136]. It is interesting to note that neither peak nor mean power appear to be unusually high in young girls who are engaged in elite tennis, swimming or gymnastics training [141]. Higher values have been recorded for elite handball players and elite sprinters [133, 141], which could not be fully explained by age and body composition; however, when comparison was made with published values for less athletic girls, peak and mean power for these elite girl athletes were not substantially different.

Longitudinal data have shown that leg peak power increases with age in both boys and girls, but the increases in boys are greater than in girls. In a study of 7- to 18-year-olds, peak power was shown to increase by 273% in girls from 7 to 16 years of age, and then to plateau [142]. In comparison, boys showed increases of 375% over this period with no plateau at 16 years. Armstrong et al. [132] examined changes in leg peak power from 12 to 17 years of age and noted increases of 66% in girls, whilst boys increased peak power by 120% over the same period. The increases noted by Armstrong et al. [132] are similar in magnitude to those of Martin et al. [142] when the same age range is considered. Similar age-related increases in mean power have been noted, again with increases in boys almost double those of girls between the ages of 12 and 17 years [132]. Sex differences in peak leg power do not appear to emerge until about 14 years of age [141], whilst mean power is greater in boys than girls from about 13 years of age [132]. Clearly age is an important predictor of short-term power in young people. The influence of stature and mass as predictors of peak and mean power have also been

established [132, 143], highlighting the need to consider both body mass and composition when assessing short-term power. De Ste Croix and colleagues [143] have shown that in addition to the effects of body mass, sum of skinfolds and age, MRI determined thigh muscle volume exerts considerable influence on young people's short-term power output. Furthermore, De Ste Croix and colleagues [143] have shown using multi-level modelling that in addition to the effects of body mass, sum of skinfolds and age, MRI-determined thigh muscle volume has a significant impact on young people's short-term power output during cycling.

There are very few data on skeletal muscle metabolism during short-duration high-intensity exercise in girls. A MRS study of pH and P_i/PCr ratio during supramaximal plantar flexion exercise in pre-pubertal and pubertal girls found that the maturational differences in pH and P_i/PCr values were not statistically significant [144]. The authors concluded that glycolytic metabolism was not maturity dependent; rather, it was dependent on muscle cross-sectional area. A more recent study of the PCr kinetics and intracellular pH response during high-intensity exercise also showed a non significant sex difference in pH. It is worthy of note that in both studies [144, 145] there was considerable variability within small samples which may be masking biological significance. Wilcox et al. [145] did demonstrate that the PCr cost per watt was higher in the girls compared to the boys [145]. These findings suggest lower efficiency in the girls compared to the boys, which may be an outcome of differences in muscle fibre type, muscle activation patterns or leg vasodilatory response. However, this study failed to demonstrate that the differences in the slow component of the PCr response between children and adults were statistically significant, raising doubt that age-related change in muscle fibre recruitment substantially influences skeletal muscle metabolism during high-intensity exercise.

Does Intensive Training Pose a Threat to the Development and Health of Young Girls?

Growth

Many young athletes begin formal training before 10 years of age, with young elite gymnasts, swimmers and tennis players entering their respective sport between the ages of 6 and 7.5 years [146]. In a number of countries young girls are recruited into specialised sport schools as young as 5 years of age [147]. These elite young athletes train intensively all year round, for many hours, with weekly training volumes of 24 h not being unusual [148]. Whether intensive training such as this distorts normal growth and maturation remains a topic of much debate [7, 149].

Evidence of reduced or delayed growth in some young athletes, such as gymnasts, has been suggested to be a direct outcome of the intensive training these youngsters have endured [5, 50, 151]. Counter-argument contends that growth reductions or delay in young athletes simply reflect their late maturation [136, 152–154]. The Training of Young Athletes (TOYA) study found that elite young female swimmers and tennis players were generally taller than the general population throughout the growth period (close to the 75th percentile for stature), whilst gymnasts were generally smaller (below the 50th percentile for stature) from 10 to 17 years of age [146]. What was noteworthy was that by 18 years of age, the gymnasts were above the 50th percentile for stature and when aligned by biological age (years from attainment of menarche), gymnasts, swimmers and tennis players showed no significant differences in height. The catch-up growth noted in the gymnasts was indicative of late maturation and apparent in girls who are not involved in competitive training, but who mature late. Both the fathers and mothers of gymnasts have been found to be significantly shorter than the parents of other athletes and genetic predisposition for stature has been not only been shown to be preserved, but often

exceeded [146, 155]. Combined, this evidence indicates that the tendency for short stature in gymnasts is not, as argued by some [151], evidence of a training-induced alteration in growth, but more likely a reflection of a genetic predisposition for later development and short stature.

Reproductive Health

Menstrual dysfunction in young athletes has also been interpreted as evidence that intensive training in young girls is harmful to reproductive health [156]. Menstrual dysfunction includes delayed menarche (onset after 16 years), luteal phase defects, oligomenorrhea and amenorrhea (table 2). Several studies have concluded that female gymnasts, swimmers and ballet dancers have delayed menarche. De Ridder et al. [29] observed that, in comparison to a control group of girls matched for maturation and fatness, girls involved in competitive gymnastics exhibited delayed menarche. An early hypothesis suggested that because these young girls had low levels of body fat they did not attain a critical level of body fat (22%) necessary for menstruation. The wide variability noted in body fat at menarche [35] has provided proof that a threshold of 22% body fat is incorrect. Additionally, there is sufficient experimental evidence in women to show that it is not body fat but caloric deprivation that affects reproductive health [157].

Genetic predisposition for late menarche in athletes has also been explored. When age of menarche in a group of elite gymnasts was correlated with maternal menarchal age, it was, on average, in lag by 0.81 years [158]. This lag was double that noted for elite swimmers and triple that for elite tennis players. These data suggest that despite a genetic predisposition for delayed menarche, this does not fully explain the extent of the delay, signalling that training may indeed delay menarche in gymnasts. However, Baxter-Jones et al. [158] went on to show that when the time period between menarchal age and retirement from the sport were

Table 2. Components of the female athlete triad, diagnosis and prevention

Component of the triad	Diagnosis	Warning signs	Prevention
Energy availability [160]	Energy availability is defined as energy intake minus exercise energy expenditure, with a threshold of $30 \text{ kcal} \cdot \text{kg}^{-1} \text{ LBM} \cdot \text{day}^{-1}$.	Low body mass (>85% of ideal body mass for stature). Fatigue.	Monitor dietary intake. Monitor training volume. Focus on healthy eating and caloric balance. Educate youngsters about nutritional fads. Reinforce message that body mass is only one aspect of good performance. Educational information on nutrition and energy expenditure, e.g. http://kidshealth.org/teen/food/sports/triad.html
Eating disorders [173]	<i>Anorexia nervosa</i> Refusal to maintain body mass over a minimally normal mass for age and stature. Intense fear of gaining mass or becoming fat. Disturbed body image. Secondary amenorrhea. <i>Bulimia nervosa</i> Recurrent episodes of binge eating (eating a large amount of food in a discrete period of time and lacking control over eating during the episode). Recurrent inappropriate compensatory behaviour such as self-induced vomiting, laxatives or excessive exercise. The binge-eating and purging behaviours occur at least twice a week for 3 months.	<i>Anorexia</i> Dramatic loss in body mass. Preoccupation with food, calories and body mass. Wears baggy clothes. Fine, downy facial hair. Mood swings. Avoidance of food-related social activities. <i>Bulimia</i> Noticeable loss in body mass. Excessive worry over weight. Bathroom visits after eating. Depression. Strict dieting followed by bingeing. Dental erosion.	Promote healthy body image. Removal of body mass/fat monitoring by coaches. Provide opportunities for developing self-coping strategies. Deemphasize body mass and thinness. Provide opportunities for nutritional counselling.
Menstrual dysfunction [174]	Oligomenorrhea – irregular menses (length between cycles >35 days). Primary amenorrhea – absence of menstruation by 15 years in girls with secondary sexual characteristics. Secondary amenorrhea- absence of menstrual cycles for 3 cycles after onset of menses.	Irregular or absent menstrual cycle.	Ask athletes to keep a training diary and include monitoring of menstrual cycle. Help girls understand that secondary amenorrhea is not normal. Provide education on reproductive health and the link between menstruation and bone health. Provide dietary education and help girls understand the link between diet and reproductive health.

Table 2. Continued

Component of the triad	Diagnosis	Warning signs	Prevention
	Luteal phase dysfunction – shortened secretory phase of the menstrual cycle, typically less than 10 days.		
Low bone mineral density [175]	If comparison with age, gender, stature and race specific Z-scores yields values ≤ 2.0 this is classified as a low bone mineral density for chronological age. Osteoporosis is diagnosed if low BMD for chronological age is accompanied by one or more of the following fracture histories: long bone fracture of the lower extremities; vertebral compression fracture and two or more long-bone fractures of the upper extremities.	Secondary amenorrhea. Stress fracture. History of fractures.	Provide educational information on osteoporosis. Provide information on nutrition for bone health, particularly focusing on calcium-rich foods. Monitor diet and provide opportunities for nutritional counselling.

considered, 92% of the girls began menarche prior to retiring. The authors concluded that training was therefore unlikely to cause the delay in menarche noted in the gymnasts, instead there was simply a chronological age difference in the timing of events. It would appear that exercise training, without other predisposing factors, is unlikely to be the cause of menstrual dysfunction.

The Female Athlete Triad

The female athlete triad was established in the early 1990s as a syndrome of three separate, but inter-related conditions, namely menstrual dysfunction, disordered eating and premature osteoporosis [159]. An updated position statement from the American College of Sports Medicine (ACSM) has revised the definition of the triad as the presence of one or more of (1) low energy

availability (with or without eating disorders), (2) amenorrhea, and (3) osteoporosis (table 2) [160]. Prevalence estimates of components of the female triad are very dependent on the athletic group studied, with higher rates in sports where low body mass is the norm. For instance, 25% of young women in endurance, weight class or aesthetic sports had clinical eating disorders, compared to 9% of the general population [161]. Secondary amenorrhea has been reported to be as high as 69% in dancers and less than 1% in the general population [162, 163]. Osteoporosis has been found in about 13% of female athletes, although this is not too different from the normal population [164] and in pre-menopausal women low bone mineral density for age is a more appropriate marker than osteoporosis.

Much of the available data on the female triad is on college-age or young adult athletes, with few reports targeting adolescents. Two key studies of

high school athletes have shown a considerable number of girls present with components of the female triad. In a study of 170 13- to 18-year-olds, 18% had disordered eating, 24% had oligomenorrhea or amenorrhea and 22% had low bone mass for their age based on WHO diagnostic criteria [165]. A higher rate of occurrence of low energy availability (55%) was noted in a study of 80 similarly aged young athletes, with 16% diagnosed with amenorrhea and using the same WHO diagnostic criteria, 16% presented with low bone mineral density [166]. The existence of the female athlete triad in these young girls is particularly worrying given this is a time when substantial amounts of bone should be accrued.

Energy deficiency appears to be particularly harmful when combined with excessive exercise, and leads to reduced oestrogen levels, athletic amenorrhea and bone demineralisation. Loucks et al. [167] have shown that there is an energy availability threshold of $20\text{--}25 \text{ kcal} \cdot \text{kg}^{-1} \text{ LBM} \cdot \text{day}^{-1}$, below which skeletal and reproductive health is compromised. This group conducted a number of studies in which women underwent energy availability manipulations, decreasing energy availability from 45 to 20 $\text{kcal} \cdot \text{kg}^{-1} \text{ LBM} \cdot \text{day}^{-1}$ or from 45 to 10 $\text{kcal} \cdot \text{kg}^{-1} \text{ LBM} \cdot \text{day}^{-1}$. These reductions caused blunting of LH pulsatility [168] and a de-linking of bone resorption and formation [169]. The existence of an energy availability threshold may help to explain why not all athletes develop athletic amenorrhea even when following the same training programme and provides a useful marker for nutritional health.

Recent conjecture that the triad is a ‘myth’ [8] has caused intense debate [164, 170, 171]. This contention stems from a number of criticisms, including flaws in the epidemiological evidence, assumptions that low energy availability implies disordered eating and a lack of experimental evidence in athletes. Much of the epidemiological evidence of the prevalence of the female triad is for individual components of the triad, rather

than for the synchronous appearance of all three which, it has been argued, over-inflates the extent of the problems. When occurrence of all three components is examined, prevalence falls dramatically [172]. On the other hand, the definition of the female athlete triad states explicitly that presence of one component is sufficient for diagnosis and the revised guidelines provided by the ACSM, have removed disordered eating and replaced it with energy availability, accepting that low energy availability does not equate to a pathological eating disorder [160].

Definitive conclusions on whether elite participation causes aberrations to skeletal and reproductive health are not possible. Nevertheless, there is clear evidence of a boundary between healthy and unhealthy levels of exertion when coupled with caloric limitation. Exposure to excessive training and caloric limitation causes abnormality in skeletal and reproductive function and regardless of the magnitude of the problem, young girl athletes deserve protection. Protection likely entails athlete education, coach recognition of the triad and the monitoring of both training volume and nutritional health in young elite girls. Table 2 provides an overview of the components of the triad, with possible prevention strategies for young athletes, coaches and parents.

Conclusions

Girls have differential growth and development in comparison to boys, resulting in substantial differences in body size and composition. Whilst stronger and leaner than many of her non-athletic peers, the smaller stature, shorter legs, lower muscularity and greater relative fatness of the elite girl athlete means she is not as strong, nor as fast as her male counterpart. Some discordant responses to exercise are not solely explained by body size and/or composition and there is evidence of underlying qualitative differences which require further clarification.

A young girl's involvement in elite sport predisposes her to increased risk of skeletal and reproductive health problems, particularly in sports where intense training is coupled with the need for leanness. Ensuring girls involved in elite training are in an environment which optimises their athletic potential while minimising risk is a priority. This entails sufficient knowledge of the physiologic responses to exercise in girls, as well as thorough understanding of the female triad disorder, its aetiology and prevention. Despite the burgeoning literature in the field of paediatric exercise physiology, an integrative understanding of girls' physiological responses to exercise remains elusive. Traditional technologies have proved

inadequate in providing a detailed understanding of the complexity of the cellular metabolic response to exercise and this is compounded by the need to separate qualitative changes from changes which are an artefact of a growing, maturing body. More recent application of non-invasive imaging techniques and breath-by-breath gas analysis is facilitating a more integrated understanding of the responses to exercise, but the number of studies to date with girls is woefully small. In addition, more needs to be learnt about the female triad and its antecedents in younger girls. The dearth of information on girls and emergence and availability of new technologies provides plenty of scope for future studies in paediatric exercise physiology.

References

- ▶ 1 Armstrong N, Welsman JR, Kirby BJ: Longitudinal changes in 11–13-year-olds' physical activity. *Acta Paediatr* 2000;89:775–780.
- 2 Women's Sport and Fitness Foundation: Participation. Factsheet November 2009:http://www.wsff.org.uk/documents/Participation_factsheet_November_09.pdf
- 3 Women's Sport and Fitness Foundation: WSSF viewpoint – women and the Olympics. Factsheet July 2009:http://www.wsff.org.uk/documents/Women_and_the_Olympics.pdf
- 4 Vertinsky P: *The Eternally Wounded Woman: Women, Doctors and Exercise in the Late Nineteenth Century*. Chicago, University of Illinois Press, 1994.
- ▶ 5 Theintz G, Howald H, Weiss U, Sizonenko P: Evidence for a reduction of growth potential in adolescent female gymnasts. *J Pediatr* 1993;122:306–313.
- ▶ 6 Damsgaard R, Bencke J, Matthiesen G, Petersen J, Müller J: Is prepubertal growth adversely affected by sport? *Med Sci Sports Exerc* 2000;32:1698–1703.
- ▶ 7 Baxter-Jones ADG, Maffulli N, Mirwald RL: Does elite competition inhibit growth and delay maturation in some gymnasts? Probably not. *Pediatr Exerc Sci* 2003;15:373–382.
- ▶ 8 DiPietro L, Stachenfeld NS: The myth of the female athlete triad. *Br J Sports Med* 2006;40:490–493.
- 9 McManus AM, Armstrong N: The elite child athlete; in Armstrong N, van Mechelen W (eds): *Paediatric Exercise Science and Medicine*, ed 2. Oxford, Oxford University Press, 2008, pp 489–502.
- 10 USA Track and Field Age Group Records. <http://www.dyestat.com/rivals/227840.html#GIRLS>
- 11 World Junior Records <http://www.iaaf.org/statistics/index.html>
- ▶ 12 Coast JR, Blevins JS, Wilson BA: Do gender differences in running performance disappear with distance? *Can J Appl Physiol* 2004;29:139–145.
- ▶ 13 George K, Sharma S, Batterham A, Whyte G, McKenna W: Allometric analysis of the association between cardiac dimensions and body size variables in 464 junior athletes. *Clin Sci* 2001;100:47–54.
- ▶ 14 Nevill AM, Holder RL: Scaling, normalizing, and per ratio standards: an allometric modeling approach. *J Appl Physiol* 1995;79:1027–1031.
- ▶ 15 Ao A, Erickson RP, Winston RML, Handyside A: Transcription of paternal Y-linked genes in the human zygote as early as the pro-nucleate stage. *Zygote* 1994;2:281–287.
- ▶ 16 Turgeon JL, Carr MC, Maki PM, Mendelsohn ME, Wise PM: Complex actions of sex steroids in adipose tissue, the cardiovascular system and brain: insights from basic science and clinical studies. *Endocr Rev* 2006;27:575–605.
- ▶ 17 Wells JCK: Sexual dimorphism of body composition. *Best Pract Res Clin Endocrinol Metab* 2007;21:415–430.
- ▶ 18 Rogol A, Clark P, Roemmich J: Growth and pubertal development in children and adolescents: effects of diet and physical activity. *Am J Clin Nutr* 2000;72:S521–S528.
- ▶ 19 Mauras N: Growth hormone and testosterone: effects on whole body metabolism and skeletal muscle in adolescence. *Horm Res* 2006;66:42–48.
- 20 Baxter-Jones ADG: Growth and maturation; in Armstrong N, van Mechelen W (eds): *Paediatric Exercise Science and Medicine*, ed 2. Oxford, Oxford University Press, 2008, pp 157–168.

- ▶ 21 Susman EJ, Houts RM, Steinberg L, Belsky J, Cauffman E, Dehart G, Friedman SL, Roisman GI, Halpern-Felsher BL, Eunice Kennedy Shriver NICHD Early Child Care Research Network: Longitudinal development of secondary sexual characteristics in girls and boys between ages 9 1/2 and 15 1/2 years. *Arch Pediatr Adolesc Med* 2010; 164:166–173.
- ▶ 22 Sherar LB, Baxter-Jones AD, Mirwald RL: Limitations to the use of secondary sex characteristics for gender comparisons. *Ann Hum Biol* 2004;31:586–593.
- 23 Malina RM, Bouchard C, Bar-Or O (eds): *Growth, Maturation and Physical Activity*. Champaign, Human Kinetics, 2004, pp 39–81.
- 24 Sinclair D, Dangerfield P: *Human Growth after Birth*, ed 6. Oxford, Oxford University Press, 1988.
- ▶ 25 Claessens AL, Veer FM, Stijnen V, Lefevre J, Maes H, Steens G, Beunen G: Anthropometric characteristics of outstanding male and female gymnasts. *J Sports Sci* 1991;9:53–74.
- ▶ 26 Damsgaard R, Bencke J, Matthiesen G, Petersen J, Müller J: Body proportions, body composition and pubertal development of children in competitive sports. *Scand J Med Sci Sports* 2001;11:54–60.
- ▶ 27 Reynolds EL, Wines JV: Individual differences in physical changes associated with adolescence in girls. *Am J Dis Child* 1948;75:329–350.
- 28 Tanner JM: *Growth at Adolescence*, ed 2. Oxford, Blackwell Scientific Publications, 1962.
- ▶ 29 de Ridder C, Thijssen J, Bruning P, Van den Brande J, Zonderland M, Erich W: Body fat mass, body fat distribution, and pubertal development: a longitudinal study of physical and hormonal sexual maturation of girls. *J Clin Endocrinol Metab* 1992;75:442–446.
- ▶ 30 Kaplowitz P: Pubertal development in girls: secular trends. *Curr Opin Obstet Gynecol* 2006;18:487–491.
- ▶ 31 Biro FM, Lucky AW, Simbartl LA, Barton BA, Daniels SR, Striegel-Moore R, Kronsberg SS, Morrison JA: Pubertal maturation in girls and the relationship to anthropometric changes: pathways through puberty. *J Pediatr* 2003;142: 643–646.
- ▶ 32 Malina R: Physical growth and biological maturation of young athletes. *Exerc Sport Sci Rev* 1994;22:389–434.
- ▶ 33 Baxter-Jones ADG, Eisenmann JC, Sherar LB: Controlling for maturation in pediatric exercise science. *Pediatr Exerc Sci* 2005;17:18–30.
- ▶ 34 Klentrou P, Plyley M: Onset of puberty, menstrual frequency, and body fat in elite rhythmic gymnasts compared with normal controls. *Br J Sports Med* 2003; 37:490–494.
- ▶ 35 Sherar LB, Baxter-Jones AD, Mirwald RL: The relationship between body composition and onset of menarche. *Ann Hum Biol* 2007;34:673–677.
- ▶ 36 Matthews BL, Bennell KL, McKay HA, Khan KM, Baxter-Jones AD, Mirwald RL, Wark JD: The influence of dance training on growth and maturation of young females: a mixed longitudinal study. *Ann Hum Biol* 2006;33:342–356.
- ▶ 37 Barrack MT, Rauh MJ, Nichols JF: Cross-sectional evidence of suppressed bone mineral accrual among female adolescent runners. *J Bone Min Res* 2010;25:1850–1857.
- ▶ 38 Rusko H, Rakkila P, Karvinen E: Anaerobic threshold, skeletal muscle enzymes and fibre composition in young female cross-country skiers. *Acta Physiol Scand* 1980;108:263–268.
- ▶ 39 Thorland WG, Johnson GO, Fagot TG, Tharp GD, Hammer RW: Body composition and somatotype characteristics of Junior Olympic athletes. *Med Sci Sports Exerc* 1981;13:332–338.
- 40 Malina RM: Regional body composition: age, sex and ethnic variation; in Roche AF, Heynsfield SB, Lohman TG (eds): *Human Body Composition*. Champaign, Human Kinetics, 1996, pp 217–255.
- 41 Falk B, Dotan R: Temperature regulation and elite young athletes; in Armstrong N, McManus AM (eds): *The Elite Young Athlete*. Med Sports Sci. Basel, Karger, 2011, vol 56, pp 126–149.
- ▶ 42 Wells JCK, Williams JE, Chomtho S, Darch T, Grijalva-Eternod C, Kennedy K, Haroun D, Wilson C, Cole TJ, Fewtrll MS: Pediatric reference data for lean tissue properties: density and hydration from age 5 to 20 y. *Am J Clin Nutr* 2010;91:610–618.
- ▶ 43 Lohman TG: Assessment of body composition in children. *Pediatr Exerc Sci* 1989;1:19–30.
- ▶ 44 Rogol AD: Growth at puberty: interaction of androgens and growth hormone. *Med Sci Sports Exerc* 1994;26:767–770.
- ▶ 45 Simoneau JA, Bouchard C: Human variation in skeletal muscle fibre-type proportion and enzyme activities. *Am J Physiol* 1989;257:E567–E572.
- 46 Staron RS, Hagerman FC, Hilida RS, Murray TF, Hostler DP, Crill MT, Ragg KE, Toma K: Fiber type composition of the vastus lateralis muscle of young men and women. *J Histochem Cytochem* 2002;48:623–629.
- ▶ 47 Schrauwen-Hinderling VB, Hesslink MKC, Schrauwen P, Kooi ME: Intramyocellular lipid content in human skeletal muscle. *Obesity* 2006;14:357–367.
- ▶ 48 Roepstorff C, Steffenson CH, Madsen M, Stallknecht B, Kanstrup L, Richter RA, Riens B: Gender differences in substrate utilization during submaximal exercise in endurance-trained subjects. *Am J Physiol Endocrinol Metab* 2002;282: E435–E447.
- ▶ 49 Hunter SK, Enoka RM: Sex differences in the fatigability of arm muscles depends on absolute force during isometric contractions. *J Appl Physiol* 2001;91:2686–2694.
- ▶ 50 Jaworowski Å, Porter MM, Holmbäck AM, Downham D, Lexell J: Enzyme activities in the tibialis anterior muscle of young moderately active men and women: relationship with body composition, muscle cross-sectional area and fibre type composition. *Acta Physiol Scand* 2002;176:215–225.
- ▶ 51 Legerlotz K, Smith HK, Hing WA: Variation and reliability of ultrasonographic quantification of the architecture of the medial gastrocnemius muscle in young children. *Clin Physiol Funct Imaging* 2010;30:198–205.
- ▶ 52 De Ste Croix MBA, Armstrong N, Welsman JR, Sharp P: Longitudinal changes in isokinetic leg strength in 10–14 year olds. *Ann Hum Biol* 2002;29:50–62.
- ▶ 53 O'Brien TD, Reeves ND, Baltzopoulos V, Jones DA, Maganaris CN: In vivo measurements of muscle specific tension in adults and children. *Exp Physiol* 2009;95:202–210.
- ▶ 54 Nygaard E: Skeletal muscle fibre characteristics in young women. *Acta Physiol Scand* 1981;112:299–304.
- ▶ 55 Vogler C, Bove KE: Morphology of skeletal muscle in children. *Arch Path Lab Med* 1985;109:238–242.

- 56 Glenmark B, Hedberg G, Jansson E: Changes in muscle fibre type from adolescence to adulthood in women and men. *Acta Physiol Scand* 1992;146:251–259.
- 57 Glenmark B, Hedberg G, Kaijster L, Jansson E: Muscle strength from adolescence to adulthood: relationship to muscle fibre types. *Eur J Appl Physiol Occup Physiol* 1994;68:9–19.
- 58 Lexell J: Human aging, muscle mass, and fibre type composition. *J Gerontol* 1995; 50:11–16.
- 59 Maffulli N, King JB, Helms P: Training in elite young athletes (the training of young athletes (TOYA study): injuries, flexibility and isometric strength. *Br J Sport Med* 1994;28:123–136.
- 60 Bencke J, Damsgaard R, Saekmose A, Jørgensen P, Jørgensen K, Klausen K: Anaerobic power and muscle strength characteristics of 11 years old elite and non-elite boys and girls from gymnastics, team handball, tennis and swimming. *Scand J Med Sci Sports* 2002;12: 171–178.
- 61 Schoenau E: Bone mass increase in puberty: what makes it happen? *Horm Res* 2006;65(Suppl 2):2–10.
- 62 Kontulainen SA, Macdonald HM, McKay HA: Change in cortical bone density and its distribution differs between boys and girls during puberty. *J Clin Endocrinol Metab* 2006;91:2555–2561.
- 63 Höglér W, Blimkie CJ, Cowell CT, Kemp AF, Briody J, Wiebe P, Farpour-Lambert N, Duncan CS, Woodhead HJ: A comparison of bone geometry and cortical density at the mid-femur between prepuberty and young adulthood using magnetic resonance imaging. *Bone* 2003;33:771–778.
- 64 Schoenau E, Neu MC, Manz F: Muscle mass during childhood: relationship to skeletal development. *J Musculoskelet Neuronal Interact* 2004;4:105–108.
- 65 Schoenau E: From mechanostat theory to development of the 'Functional Muscle-Bone-Unit'. *J Musculoskelet Neuronal Interact* 2005;5:232–238.
- 66 Ward KA, Roberts SA, Adams JE, Mughal MZ: Bone geometry and density in the skeleton of pre-pubertal gymnasts and school children. *Bone* 2005;36: 1012–1018.
- 67 Duncan CS, Blimkie CJR, Cowell CT, Burke ST, Briody JN, Howman-Giles R: Bone mineral density in adolescent female athletes: relationship to exercise type and muscle strength. *Med Sci Sports Exerc* 2002;34:286–294.
- 68 Haapasalo H, Kontulainen S, Sievanen H, Kannus P, Jarvinen M, Vuori I: Exercise-induced bone gain is due to enlargement in bone size without a change in volumetric bone density: a peripheral quantitative computed tomography study of the upper arms of male tennis players. *Bone* 2000;27:351–357.
- 69 Duncan CS, Blimkie CJR, Kemp A, Higgs W, Cowell CT, Woodhead H, Briody JN, Howman-Giles R: Mid-femur geometry and biomechanical properties in 15- to 18-year-old female athletes. *Med Sci Sports Exerc* 2002;34:673–681.
- 70 Höglér W, Blimkie CJR, Cowell CT, Inglis, Rauch F, Kemp AF, Wiebe P, Duncan CS, Farpour-Lambert N, Woodhead HJ: Sex-specific developmental changes in muscle size and bone geometry at the femoral shaft. *Bone* 2008;42:982–989.
- 71 Douthwaite JN, Kanaley JA, Spadaro JA, Hickman RM, Scerpella TA: Muscle indices do not fully account for enhanced upper extremity bone mass and strength in gymnasts. *J Musculoskelet Neuronal Interact* 2009;9:2–14.
- 72 Daly RM, Saxon L, Turner CH, Robling AG, Bass SL: The relationship between muscle size and bone geometry during growth and in response to exercise. *Bone* 2004;34:281–287.
- 73 Bailey DA, Martin AD: Physical activity and skeletal health in adolescents. *Pediatr Exerc Sci* 1996;6:330–347.
- 74 Tanner JM: Fallacy of per-weight and per-surface area standards and their relation to spurious correlation. *J Appl Physiol* 1949;2:1–15.
- 75 Guenette A, Sheel AW: Exercise-induced arterial hypoxaemia in active young women. *Appl Physiol Nutr* 2007;32: 1263–1273.
- 76 Mota J, Guerra S, Leandro C, Pinto A, Ribeiro JC, Duarte JA: Association of maturation, sex, and body fat in cardio-respiratory fitness. *Am J Human Biol* 2002;14:707–712.
- 77 Armstrong N, Welsman JR: Aerobic fitness; in Armstrong N, van Mechelen W (eds): *Paediatric Exercise Science and Medicine*, ed 2. Oxford, Oxford University Press, 2008, pp 97–108.
- 78 Nevill AM, Bate S, Holder RL: Modeling physiological and anthropometric variables known to vary with body size and other confounding variables. *Year Book Phys Anthropol* 2005;48:141–153.
- 79 Tolfrey K, Barker A, Thom JM, Morse CI, Narici MV, Batterham AM: Scaling of maximal oxygen uptake by lower leg muscle volume in boys and men. *J Appl Physiol* 2006;100:1851–1856.
- 80 Rowland T, Goff D, Martel L, Ferrone L, Kline G: Normalization of maximal cardiovascular variables for body size in premenarchal girls. *Pediatr Cardiol* 2000;21:429–432.
- 81 Armstrong N, McManus AM, Welsman JR: Aerobic fitness; in Armstrong N, van Mechelen W (eds): *Paediatric Exercise Science and Medicine*, ed 2. Oxford, Oxford University Press, 2008, pp 269–282.
- 82 Armstrong N, Welsman JR: *Young People and Physical Activity*. Oxford, Oxford University Press, 1997.
- 83 Armstrong N, Welsman JR: Peak oxygen uptake in relation to growth and maturation in 11- to 17-year-old humans. *Eur J Appl Physiol* 2001;85:546–551.
- 84 Eisenmann JC, Pivarnik JM, Malina RM: Scaling peak $\dot{V}O_2$ to body mass in young male and female distance runners. *J Appl Physiol* 2001;90:2172–2180.
- 85 Baxter-Jones A, Goldstein H, Helms P: The development of aerobic power in young athletes. *J Appl Physiol* 1993;75: 1160–1167.
- 86 Welsman JR, Armstrong N: Longitudinal changes in submaximal oxygen uptake in 11- to 13-year-olds. *J Sports Sci* 2000; 18:183–189.
- 87 Courteix D, Obert P, Lecoq AM, Guenon P, Kock G: Effect of intensive swimming training on lung volumes, airway resistances and on the maximal expiratory flow-volume relationship in prepubertal girls. *Eur J Appl Physiol* 1997;76:264–269.
- 88 Sheel AW, Richards JC, Foster GE, Guenette JA: Sex differences in respiratory exercise physiology. *Sports Med* 2004;34:567–579.
- 89 Boezen HM, Jansen DF, Postma DS: Sex and gender differences in lung development and their clinical significance. *Clin Chest Med* 2004;25:237–245.
- 90 Harms CA: Does gender affect pulmonary function and exercise capacity? *Respir Physiol Neurobiol* 2006;151:124–131.

- ▶ 91 Guenette A, Sheel AW: Exercise-induced arterial hypoxaemia in active young women. *Appl Physiol Nutr* 2007;32: 1263–1273.
- ▶ 92 Olfert IM, Balouch J, Kleinsasser A, Knapp A, Wagner H, Wagner PD, Hopkins SR: Does gender affect human pulmonary gas exchange during exercise? *J Physiol* 2004;557:529–541.
- 93 Fawcner SG, Armstrong N: Pulmonary function; in Armstrong N, van Mechelen W (eds): *Paediatric Exercise Science and Medicine*, ed 2. Oxford, Oxford University Press, 2008, pp 243–254.
- ▶ 94 Armstrong N, Kirby BJ, McManus AM, Welsman JR: Prepubescents' ventilatory responses to exercise with reference to sex and body size. *Chest* 1997;112:1554–1560.
- ▶ 95 Rowland TW, Cunningham LN: Development of ventilatory responses to exercise in normal white children. *Chest* 1997;111:327–332.
- ▶ 96 Hopkins SR, Harms CA: Gender and pulmonary gas exchange during exercise. *Exerc Sport Sci Rev* 2004;32:50–56.
- ▶ 97 Swain KE, Rosenkranz SK, Beckman B, Harms CA: Expiratory flow limitation during exercise in prepubescent boys and girls: prevalence and implications. *J Appl Physiol* 2010;108:1267–1274.
- ▶ 98 Laursen PB, Tsang GCK, Smith GJ, van Velzen MV, Ignatova BB, Sprules EB, Chu KS, Coutts KD, McKenzie DC: Incidence of exercise-induced arterial hypoxemia in prepubescent females. *Pediatr Pulmonol* 2002;34:37–41.
- ▶ 99 Guerro L, Naranjo J, Carranza MD: Influence of gender on ventilatory efficiency during exercise in young children. *J Sports Sci* 2008;26:1455–1457.
- ▶ 100 Karlberg P, Lind J: Studies of the total amount of hemoglobin and the blood volume in children. 1. Determination of total hemoglobin and blood volume in normal children. *Acta Paediatr Scand* 1955;44:17–34.
- ▶ 101 Sukarochana K, Parenzan L, Thakurdas N, Kiesewetter WB: Red cell mass determinations in infancy and childhood, with the use of radioactive chromium. *J Pediatr* 1961;59:903–908.
- ▶ 102 Linderkamp O, Versmold HT, Riegel KP, Betke K: Estimation and prediction of blood volume in infants and children. *Eur J Pediatr* 1977;125:227–234.
- ▶ 103 Raes A, Van Aken S, Craen M, Donckerwolcke R, Walle JV: A reference frame for blood volume in children and adolescents. *BMC Pediatrics* 2006;6:3.
- ▶ 104 Dallman PR, Siimes MA: Percentile curves for hemoglobin and red cell volume in infancy and childhood. *Pediatrics* 1979;94:26–31.
- ▶ 105 Boyadjiev N, Taralov Z: Red blood cell variables in highly trained pubescent athletes: a comparative analysis. *Br J Sports Med* 2000;34:200–204.
- 106 Rowland TW: *Developmental Exercise Physiology*. Champaign, Human Kinetics, 1996, pp 117–140.
- ▶ 107 Larsen JA, Kadish AH: Effects of gender on cardiac arrhythmias. *J Cardiovasc Electrophysiol* 1998;9:655–664.
- ▶ 108 Wirth A, Trager E, Scheele K, Mayer D, Diehm K, Reischle K, Weicker H: Cardiopulmonary adjustment and metabolic response to maximal and submaximal physical exercise of boys and girls at different stages of maturity. *Eur J Appl Physiol* 1978;39:29–39.
- ▶ 109 Telford RD, McDonal IG, Ellis LB, Chennells MH, Sandstrom ER, Fuller PJ: Echocardiographic dimensions in trained and untrained 12-year-old-boys and girls. *J Sports Sci* 1988;6:49–57.
- 110 Thoren CAR, Asano K: Functional capacity and cardiac function in 10-year-old-boys and girls with high and low running performance; in Ilmarinen J, Valimaki I (eds): *Children and Sport: Pediatric Work Physiology*. Berlin, Springer, 1984 pp 182–188.
- ▶ 111 Cumming GR: Hemodynamics of supine bicycle exercise in 'normal' children. *Am Heart J* 1977;93:617–622.
- ▶ 112 Rowland TW, Goff D, Martel L, Ferrone L: Influence of cardiac functional capacity on gender differences in maximal oxygen uptake in children. *Chest* 2000; 117:629–635.
- ▶ 113 Obert P, Mandigout S, Nottin S, Vinet A, N'Guyen L, Lecoq A: Cardiovascular responses to endurance training in children: effect of gender. *Eur J Clin Invest* 2003;33:199–208.
- ▶ 114 Winsley RJ, Fulford J, Roberts AC, Welsman JR, Armstrong N: Sex differences in peak oxygen uptake in prepubertal children. *J Sci Med Sport* 2009;12: 647–651.
- 115 Gutin B, Owens S, Trieber F, Mensah G: Exercise hemodynamics and left ventricular parameters in children; in Armstrong N, Kirby B, Welsman J (eds): *Children and Exercise*. Part XIX. London, E & FN Spon, 1997, pp 460–464.
- ▶ 116 Osika W, Dangardt F, Montgomery SM, Volkmann R, Li MG, Friberg P: Sex differences in peripheral artery intima, media and intima media thickness in children and adolescents. *Atherosclerosis* 2009;203:172–177.
- ▶ 117 Gavin KM, Seals DR, Silver AE, Moreau KL: Vascular endothelial estrogen receptor alpha is modulated by estrogen status and related to endothelial function and endothelial nitric oxide synthase in healthy women. *J Clin Endocrinol Metab* 2009;94:3513–3520.
- ▶ 118 Nadland IH, Wallocl L, Toska K: Effect of the leg muscle pump on the rise in muscle perfusion during muscle work in humans. *Eur J Appl Physiol* 2009;105: 829–841.
- ▶ 119 Rowland T, Lisowski R: Hemodynamic responses to increasing cycle cadence in 11-year old boys: role of the skeletal muscle pump. *Int J Sports Med* 2001;22: 405–409.
- ▶ 120 Armstrong N, Welsman JR: Cardiovascular responses to submaximal treadmill running in 11 to 13 year olds. *Acta Paediatr* 2002;91:125–131.
- 121 Constantin-Theodosiu D, Greenhaff PL, McIntyre DB, Round JM, Jones DA: Anaerobic energy production in human skeletal muscle in intense contraction: a comparison of ³¹P magnetic resonance spectroscopy and biochemical techniques. *Exp Physiol* 1997;82:592–601.
- ▶ 122 Barker A, Welsman J, Welford D, Fulford J, Williams C, Armstrong N: Reliability of ³¹P magnetic resonance spectroscopy during an exhaustive exercise test in children. *Eur J Appl Physiol* 2006;98: 556–565.
- 123 Armstrong N, Fawcner SG: Exercise metabolism; in Armstrong N, van Mechelen W (eds): *Paediatric Exercise Science and Medicine*, ed 2. Oxford, Oxford University Press, 2008, pp 213–226.
- ▶ 124 Jones AM, Poole DC: Oxygen uptake dynamics: from muscle to mouth—an introduction to the symposium. *Med Sci Sports Exerc* 2005;37:1542–1550.
- ▶ 125 Barker AR, Welsman JR, Fulford J, Welford D, Williams CA, Armstrong N: Muscle phosphocreatine and pulmonary oxygen uptake kinetics in children at the onset and offset of moderate intensity exercise. *Eur J Appl Physiol* 2008;102: 727–738.
- ▶ 126 Fawcner SG, Armstrong N: Oxygen uptake kinetic response to exercise in children. *Sports Med* 2003;33:651–669.

- ▶ 127 Fawcner SG, Armstrong N, Potter CR, Welsman JR: Oxygen uptake kinetics in children and adults after the onset of moderate-intensity exercise. *J Sports Sci* 2002;20:319–326.
- ▶ 128 Barker AR, Welsman JR, Fulford J, Welford D, Armstrong N: Muscle phosphocreatine kinetics in children and adults at the onset and offset of moderate-intensity exercise. *J Appl Physiol* 2008;105:446–456.
- ▶ 129 Fawcner SG, Armstrong N: Longitudinal changes in the kinetic response to heavy intensity exercise. *J Appl Physiol* 2004; 97:460–466.
- ▶ 130 Barker AR, Welsman JR, Fulford J, Welford D, Armstrong N: Quadriceps muscle energetic during incremental exercise in children and adults. *Med Sci Sports Exerc* 2011;42:1303–1313.
- ▶ 131 Rowland TW: Circulatory responses to exercise: are we misleading Fick? *Chest* 2005;127:1023–1030.
- ▶ 132 Armstrong N, Welsman JR, Williams CA, Kirby BJ: Longitudinal changes in young people's short-term power output. *Med Sci Sports Exerc* 2000;32:1140–1145.
- ▶ 133 Thorland WG, Johnson GO, Cisar CJ, Housh TJ, Tharp GD: Strength and anaerobic responses of elite young female sprint and distance runners. *Med Sci Sports Exerc* 1987;19:56–61.
- 134 Blimkie CJ, Roche P, Bar-Or O: The anaerobic power ratio in adolescent boys and girls; in Rutenfranz J, Mocellin R, Klimt F (eds): *Children and Exercise XII*. International Series on Sport Science. Champaign, Human Kinetics, 1986, pp 31–37.
- ▶ 135 Palgi Y, Gutin B, Young J, Alejandro D: Physiological and anthropometric factors underlying endurance performance in children. *Int J Sports Med* 1984;5:67–73.
- 136 Bar-Or O: *Pediatric Sports Medicine for the Practitioner (Comprehensive Manual in Pediatrics)*. New York, Springer, 1983.
- ▶ 137 Doré E, Bedu M, França NM, Van Praagh E: Anaerobic cycling performance characteristics in prepubescent, adolescent and young adult females. *Eur J Appl Physiol* 2001;84:476–481.
- ▶ 138 Aucouturier J, Lazaar N, Doré E, Meyer M, Ratel S, Duché P: Cycling peak power in obese and lean 6–8 year-old girls and boys. *Appl Physiol Nutr Metab* 2007;32: 367–371.
- ▶ 139 Van Praagh E: Development of anaerobic function during childhood and adolescence. *Pediatr Exerc Sci* 1990;2:336–348.
- 140 Williams CA, Armstrong N: Optimised peak power output of adolescent children during maximal sprint pedalling; in Ring FJ (ed): *Children in Sport*. Proc 1st Bath Sports Medicine Conference. Bath, University of Bath, 1995, pp 40–44.
- ▶ 141 Bencke J, Damsgaard R, Saekmose A, Jørgensen P, Jørgensen K, Klausen K: Anaerobic power and muscle strength characteristics of 11 years old elite and non-elite boys and girls from gymnastics, team handball, tennis and swimming. *Scand J Med Sci Sports* 2002;12: 171–178.
- ▶ 142 Martin RJF, Doré E, Twisk J, Van Praagh E, Hautier CA, Bedu M: Longitudinal changes of maximal short-term peak power in girls and boys during growth. *Med Sci Sports Exerc* 2004;36:498–503.
- ▶ 143 De Ste Croix MBA, Armstrong N, Chia MYH, Welsman JR, Parsons G, Sharpe P: Changes in short-term power output in 10- to 12-year-olds. *J Sport Sci* 2001;19: 141–148.
- ▶ 144 Peterson SR, Gaul CA, Stanton MM, Hanstock CC: Skeletal muscle metabolism during short-term, high-intensity exercise in prepubertal and pubertal girls. *J Appl Physiol* 1999;87:2151–2156.
- ▶ 145 Wilcocks RJ, Williams CA, Barker AR, Fulford J, Armstrong N: Age- and sex-related differences in muscle phosphocreatine and oxygenation kinetics during high-intensity exercise in adolescents and adults. *NMR Biomed* 2010; 3:569–577.
- ▶ 146 Erlandson MC, Sherar LB, Mirwald RL, Maffulli N, Baxter-Jones AD: Growth and maturation of adolescent female gymnasts, swimmers, and tennis players. *Med Sci Sports Exerc* 2008;40:34–42.
- ▶ 147 Hong F: Innocence lost: child athletes in China. *Sport Society* 2004;7:338–354.
- ▶ 148 Lang M: Surveillance and conformity in competitive youth swimming. *Sport Educ Society* 2010;15:19–37.
- ▶ 149 Caine D, Bass S, Daly R: Does elite competition inhibit growth and delay maturation in some gymnasts? Quite possibly. *Pediatr Exerc Sci* 2003;15:360–372.
- ▶ 150 Klentrou P, Plyley M: Onset of puberty, menstrual frequency, and body fat in elite rhythmic gymnasts compared with normal controls. *Br J Sports Med* 2003;37:490–494.
- 151 Bass S, Bradney G, Pearce E, Hendrich E, Inge K, Stuckey S, Lo SK, Seeman E: Short stature and delayed puberty in gymnasts: influence of selection bias on leg length and the duration of training on trunk length. *J Pediatr* 2000;1136: 149–155.
- ▶ 152 Georgopoulos N, Markou K, Theodoropoulou A, Paraskevopoulou P, Varaki L, Kazantzi Z, Leglise M, Vagenakis AG: Growth and pubertal development in elite female rhythmic gymnasts. *J Clin Endocrinol Metab* 1999;84:4525–4530.
- ▶ 153 Georgopoulos NA, Markou KB, Theodoropoulou A, Vagenakis GA, Benardot D, Leglise M, Dikmopoulos JCA, Vagenakis AP: Height velocity and skeletal maturation in elite female rhythmic gymnasts. *J Clin Endocrinol Metab* 2001;86:5159–5164.
- ▶ 154 Theodoropoulou A, Markou KB, Vagenakis GA, Benardot D, Leglise M, Kourounis G, Vagenakis AP, Georgopoulos NA: Delayed but normally progressed puberty is more pronounced in artistic compared with rhythmic elite gymnasts due to the intensity of training. *J Clin Endocrinol Metab* 2005; 90:6022–6027.
- ▶ 155 Baxter-Jones ADG, Helms PJ: Effects of training at a young age: a review of the Training of Young Athletes (TOYA) Study. *Pediatr Exerc Sci* 1996;8:310–327.
- ▶ 156 Constantini NW, Warren MP: Menstrual dysfunction in swimmers: a distinct entity. *J Clin Endocrinol Metab* 1995;80: 2740–2744.
- ▶ 157 Loucks AB: Energy availability, not body fatness, regulates reproductive function in women. *Exerc Sport Sci Rev* 2003;31: 144–148.
- ▶ 158 Baxter-Jones ADG, Helms P, Baines-Preece J, Preece M: Menarche in intensively training gymnasts, swimmers and tennis players. *Ann Hum Biol* 1994;21: 407–415.
- 159 Otis CL, Drinkwater B, Johnson M, Loucks A, Wilmore J: American College of Sports Medicine position stand: the female athlete triad. *Med Sci Sports Exerc* 1997;29:i–ix.
- ▶ 160 Nattiv A, Loucks AB, Manore MM, Sanborn CF, Sundgot-Borgen J, Warren MP, American College of Sports Medicine: American College of Sports Medicine position stand: the female athlete triad. *Med Sci Sports Exerc* 2007;39: 1867–1882.

- ▶ 161 Sundgot-Borgen J, Torstveit MK: Prevalence of eating disorders in elite athletes is higher than in the general population. *Clin J Sport Med* 2004;14: 25–32.
- ▶ 162 Abraham SF, Beumont PJ, Fraser IS, Llewellyn-Jones D: Body weight, exercise and menstrual status among ballet dancers in training. *Br J Obstet Gynaecol* 1982;89:507–510.
- ▶ 163 Chumlea WC, Schubert CM, Roche AF, Kulin HE, Lee PA, Himes JH, Sun SS: Age at menarche and racial comparisons in US girls. *Pediatrics* 2003;111:110–113.
- ▶ 164 Beals, KA, Meyer NL: Female athlete triad update. *Clin Sports Med* 2007;26: 69–89.
- ▶ 165 Nichols JE, Rauh MJ, Lawson MJ, Ming JJ, Barkai HS: Prevalence of the female athlete triad syndrome among high school athletes. *Arch Pediatr Adolesc Med* 2006;160:137–142.
- ▶ 166 Hoch AZ, Pajewski NM, Moraski L, Carrera GF, Wilson CR, Hoffmann RG, Schimke JE, Gutterman DD: Prevalence of the female athlete triad in high school athletes and sedentary students. *Clin J Sport Med* 2009;19:421–428.
- 167 Loucks AB, Heath EM: Induction of low T₃ syndrome in exercising women occurs at a threshold of energy availability. *Am J Physiol* 1994;226:R817–R823.
- ▶ 168 Loucks AB, Thurma JA: Luteinizing hormone pulsatility is disrupted at a threshold of energy availability in regularly menstruating women. *J Clin Endocrinol Metab* 2003;88:297–311.
- ▶ 169 Ihle R, Loucks AB: Dose-response relationships between energy availability and bone turnover in young exercising women. *J Bone Miner Res* 2004;19: 1231–1240.
- ▶ 170 Loucks AB: Refutation of 'the myth of the female athlete triad'. *Br J Sports Med* 2007;41:55–57.
- ▶ 171 De Souza MJ, Vescovi JD, Willijmas NI, Van Heest JL, Warren MP: Correction of misinterpretation and misrepresentations of the female athlete triad. *Br J Sports Med* 2007;41:58–59.
- ▶ 172 Klungland-Torstveig M, Sundgot-Borgen J: The female athlete triad exists in both elite athletes and controls. *Med Sci Sports Exerc* 2005;37:1449–1459.
- 173 American Psychiatric Association: Eating disorders; in First M (ed): *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. Washington, American Psychiatric Publishing, 1994 pp 539–550.
- ▶ 174 Practice Committee of the American Society for Reproductive Medicine: Current evaluation of amenorrhea. *Fertil Steril* 2004;82:266–272.
- ▶ 175 Bianchi ML, Baim S, Bishop NJ, Gordon CM, Hans DB, Langman CB, Leonard MB, Kalkwarf HJ: Official positions of the International Society for Clinical Densitometry (ISCD) on DXA evaluation in children and adolescents. *Pediatr Nephrol* 2010;25:37–47.

Dr. Alison M. McManus
 Institute of Human Performance, University of Hong Kong
 Hong Kong, SAR (China)
 Tel. +852 2589 0582, Fax +852 2855 1712, E-Mail alimac@hku.hk

TRANSGENDER HEALTH

Early Hormonal Treatment Affects Body Composition and Body Shape in Young Transgender Adolescents



Maartje Klaver, MD,¹ Renée de Mutsert, PhD,² Chantal M. Wiepjes, MD,¹ Jos W. R. Twisk, PhD,³ Martin den Heijer, MD, PhD,¹ Joost Rotteveel, MD, PhD,⁴ and Daniël T. Klink, MD, PhD^{4,5}

ABSTRACT

Background: Transgender adolescents aspiring to have the body characteristics of the affirmed sex can receive hormonal treatment. However, it is unknown how body shape and composition develop during treatment and whether transgender persons obtain the desired body phenotype.

Aim: To examine the change in body shape and composition from the start of treatment with gonadotropin-releasing hormone agonists (GnRHa) until 22 years of age and to compare these measurements at 22 years with those of age-matched peers.

Methods: 71 transwomen (birth-assigned boys) and 121 transmen (birth-assigned girls) who started treatment from 1998 through 2014 were included in this retrospective study. GnRHa treatment was started and cross-sex hormonal treatment was added at 16 years of age. Anthropometric and whole-body dual-energy x-ray absorptiometry data were retrieved from medical records. Linear mixed model regression was performed to examine changes over time. SD scores (SDS) were calculated to compare body shape and composition with those of age-matched peers.

Outcomes: Change in waist-hip ratio (WHR), total body fat (TBF), and total lean body mass (LBM) during hormonal treatment. SDS of measures of body shape and composition compared with age-matched peers at 22 years of age.

Results: In transwomen, TBF increased (+10%, 95% CI = 7–11) while total LBM (–10%, 95% CI = –11 to –7) and WHR (–0.04, 95% CI = –0.05 to –0.02) decreased. Compared with ciswomen, SDS at 22 years of age were +0.3 (95% CI = 0.0–0.5) for WHR, and 0.0 (95% CI = –0.2 to 0.3) for TBF. Compared with cismen, SDS were –1.0 (95% CI = –1.3 to –0.7) for WHR, and +2.2 (95% CI = 2.2–2.4) for TBF. In transmen, TBF decreased (–3%, 95% CI = –4 to –1), while LBM (+3%, 95% CI = 1–4) and WHR (+0.03, 95% CI = 0.01–0.04) increased. Compared with ciswomen, SDS at 22 years of age were +0.6 (95% CI = 0.4–0.8) for WHR, and –1.1 (95% CI = –1.4 to –0.9) for TBF. Compared with cismen, SDS were –0.5 (95% CI = –0.8 to –0.3) for WHR, and +1.8 (95% CI = 1.6–1.9) for TBF.

Clinical Implications: Knowing body shape and composition outcomes at 22 years of age will help care providers in counseling transgender youth on expectations of attaining the desired body phenotype.

Strengths and Limitations: This study presents the largest group of transgender adults to date who started treatment in their teens. Despite missing data, selection bias was not found.

Conclusions: During treatment, WHR and body composition changed toward the affirmed sex. At 22 years of age, transwomen compared better to age-matched ciswomen than to cismen, whereas transmen were between reference values for ciswomen and cismen. **Klaver M, de Mutsert R, Wiepjes CM, et al. Early Hormonal Treatment Affects Body Composition and Body Shape in Young Transgender Adolescents. J Sex Med 2018;15:251–260.**

Received September 20, 2017. Accepted December 14, 2017.

¹Department of Endocrinology and Center of Expertise on Gender Dysphoria, VU University Medical Center, Amsterdam, The Netherlands;

²Department of Clinical Epidemiology, Leiden University and Medical Center, Leiden, the Netherlands;

³Department of Clinical Epidemiology, VU University, Amsterdam, The Netherlands;

⁴Department of Pediatrics, Division of Endocrinology, VU University Medical Center, Amsterdam, The Netherlands;

⁵Department of Pediatrics, Division of Pediatric Endocrinology, Ghent University Hospital, Ghent, Belgium

Copyright © 2017, International Society for Sexual Medicine, Inc. All rights reserved. <https://doi.org/10.1016/j.jsxm.2017.09.001>

Copyright © 2017, International Society for Sexual Medicine. Published by Elsevier Inc. All rights reserved.

Key Words: Transgender Persons; Adolescent; Gonadotropin-Releasing Hormone Analogues; Cross-Sex Hormonal Treatment; Body Composition; Body Shape

INTRODUCTION

Adolescents with gender dysphoria aspire to have body characteristics that are similar to those of the affirmed sex. From 12 years of age, adolescents with male-to-female gender dysphoria, referred to as transwomen, and adolescents with female-to-male gender dysphoria, referred to as transmen, can be treated with gonadotropin-releasing hormone analogues (GnRHa) to suppress puberty. Subsequently, at 16 years of age and if the person still pursues gender-affirming treatment, cross-sex hormonal treatment (CHT) is added to induce the secondary sexual characteristics of the affirmed sex.¹

During puberty, with increasing sex steroid levels, girls develop more body fat that is deposited mainly in the gluteal and femoral region (so-called gynoid region).^{2,3} This leads to a female body shape with a low waist-to-hip ratio (WHR).^{3,4} Pubertal boys obtain more lean body mass (LBM) and store body fat mainly in the abdominal region (also referred to as the android region),³ resulting in a male body shape with a higher WHR than seen in girls.^{3,4} It is unknown how total and regional body fat, LBM, and body shape develop in transgender adolescents treated with GnRHa and CHT, and whether this results in a similar body composition and body shape as the affirmed sex in young adulthood.

Therefore, the 1st aim of this study was to examine the effects of treatment with GnRHa and CHT on total body and regional body fat, LBM, and body shape in adolescents with gender dysphoria. A 2nd aim of this study was to compare the achieved amount of total and regional body fat, LBM, and WHR of these transwomen and transmen at 22 years of age with reference values of the affirmed sex and to examine whether they obtained the desired body composition and body shape in young adulthood. A 3rd aim was to examine the influence of pubertal stage at start of treatment on the achieved body composition and body shape at 22 years.

METHODS

Study Design and Study Population

We retrospectively reviewed the medical records of all adolescents diagnosed with gender dysphoria (*Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision*⁵) at the VU University Medical Center (Amsterdam, the Netherlands) until December 2015. All persons who started hormonal treatment before 18 years of age, started the treatment protocol as described below,¹ had undergone whole-body dual-energy x-ray absorptiometry (DXA) during treatment, and

according to their age had their medical checkups in young adulthood (>20.5 years) were eligible for this study. Data obtained during routine medical checkups on anthropometry, laboratory measurements, and whole-body DXA were collected at 3 time points: start of GnRHa, addition of CHT, and result at 22 years of age (range = 20.5–23.5 years). The local ethics committee approved the study and the necessity for informed consent was waived.

Treatment Protocol

The treatment protocol, also referred to as the Dutch protocol, has been published in detail.¹ At a minimum age of 12 years and stage B2 (breast) for girls and Tanner stage G3 (genital) for boys, subcutaneous GnRHa 3.75 mg for 4 weeks was started. From 16 years of age, CHT was added with increasing doses to initiate pubertal development. Transwomen were prescribed oral 17 β -estradiol starting at 5 μ g per kilogram of body weight per day, which was increased by 5 μ g/kg per day every 6 months until the maintenance dose of 2 mg/day was reached. Transmen used initially mixed testosterone esters (Sustanon; Organon Pharmaceuticals, Oss, The Netherlands) intramuscularly starting at 25 mg per square meter of body surface area every 2 weeks, which was increased by 25 mg/m² every 6 months until the maintenance dose of 250 mg every 3 to 4 weeks was achieved. When GnRHa were started after 16 years of age, CHT was added after 3 to 6 months with a start dosage of 17 β -estradiol 1 mg/day or intramuscular Sustanon 75 mg/week. After 6 months, this was increased to 17 β -estradiol 2 mg/day in transwomen and Sustanon 250 mg every 3 to 4 weeks in transmen. From 18 years, patients were eligible for gonadectomy, after which treatment with GnRHa ceased. From the start of treatment, patients were advised to maintain a healthy lifestyle with sportive activities and an adequate calcium intake to prevent bone loss.

Anthropometry and Whole-Body DXA

At each visit, body height, body weight, waist circumference, and hip circumference were measured. Body height was measured to the nearest 0.1 cm using a Harpenden stadiometer. Body weight was measured while the subject wore only underwear without shoes to the nearest 0.1 kg. Body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared. Waist circumference, defined as the smallest abdominal circumference, and hip circumference, measured at the level of the trochanter major, were determined with a tape measure to the nearest 0.1 cm. From these 2 measurements, the

WHR was calculated, which was used as a measure for body shape.

Whole-body and regional body fat, LBM, and total mass were measured using DXA. Until 2002, the Hologic QDR 2000 (Hologic Inc, Bedford, MA, USA) was used. From 2002, a Hologic Delphi apparatus (Hologic Inc) was used with software version 8.26, which was updated in 2005 to version 12.3.3. In 2011, the Hologic Delphi was replaced by a Hologic Discovery 13.1 (Hologic Inc), which was updated to version 3.3 in 2012 and to version 4.5.3 in 2015. During the review of medical records, DXA scans were included when they were obtained within 4 months before or after the start of GnRHa or CHT or within 1.5 years before or after the 22nd birthday. All available DXA scans from participants were obtained and reanalyzed with the most recent software version (version 13.5.3). The android region and gynoid region were defined using the software provided by Hologic. The lower boundary of the android region coincides with the upper edge of the pelvis and the height equals 20% of the distance from the upper edge of the pelvis to the bottom of the chin. The upper boundary of the gynoid region is below the upper edge of the pelvis by 1.5 times the height of the android region. The gynoid region equals twice the height of the android region.

Statistical Analyses

STATA 13.1 (StataCorp, College Station, TX, USA) was used for statistical analyses. Linear mixed model regression analyses with observations clustered within participants were performed to examine mean changes in measurements of body composition and body shape from the start of GnRHa to 22 years of age. Linear mixed models also properly deal with missing data.⁶ When outcome variables were not normally distributed, the natural logarithm was obtained for analyses. Analyses were adjusted for Tanner stage at start of treatment, and BMI at start of treatment using centered variables. Student t-tests were used to examine whether changes in body shape and body composition during GnRHa monotherapy differed from changes after the addition of CHT.

Tanner stage at start of treatment, time, and the interaction between Tanner stage and time were added to the linear mixed model regression analysis to examine the influence on the achieved body composition and body shape at 22 years of age. In transwomen, Tanner stage at start was defined by testes volume and this resulted in the following categories: less than or equal to 8 mL (early puberty), 10 to 15 mL (mid-puberty), and at least 20 mL (late puberty).^{7,8} In transmen, Tanner stage at start of treatment was defined by breast development.⁷ Because a small number of transmen had Tanner stage II ($n = 3$) or III ($n = 8$) at the start of CHT, the 2 groups were classified as starters in early and mid-puberty. Transmen with Tanner stage IV or V were classified as starters in late puberty. Analyses were adjusted for BMI at 22 years of age.

We calculated standard deviation scores (SDS) to compare measures of body composition and body shape in participants at

22 years of age with reference values from age-matched peers with no (treatment for) gender dysphoria, also referred to as ciswomen and cismen. In transwomen and transmen, SDS were calculated for both ciswomen and cismen in order to compare measures with the at birth assigned sex and the affirmed sex. For the mean age of start of GnRHa (15 years in both transwomen and transmen) and for the age of 22 years, age-specific reference values^{4,9-12} were retrieved from literature.

RESULTS

71 transwomen and 121 transmen who started with GnRHa and CHT from 1998 through 2014 were included in the present analyses (Figure 1). The general characteristics of the participants are presented in Table 1.

Change in Body Shape and Body Composition During Treatment in Transwomen

The results of the mixed model analyses showed that in transwomen waist circumference (+8 cm, 95% CI = 5–10, $P < .001$) and hip circumference (+17 cm, 95% CI = 13–21, $P < .001$) increased and WHR decreased (−0.04, 95% CI = −0.05 to −0.02, $P < .001$). The percentage of total body fat increased (+9%, 95% CI = 8–11, $P < .001$) and thus the percentage of LBM decreased (−9%, 95% CI = −11 to −8, $P < .001$; Table 2, Figure 2). Percentage of body fat increased in the android region (+9%, 95% CI = 6–12, $P < .001$) and the gynoid region (+11%, 95% CI = 9–12, $P < .001$; Table 2). Changes in body composition and body shape measurements were not different after adjustment for Tanner stage at start of treatment, or BMI at start of treatment.

Change in Body Shape and Body Composition During Treatment in Transmen

Transmen showed increases in waist circumference (+6 cm, 95% CI = 4–8, $P < .001$), hip circumference (+5 cm, 95% CI = 2–7, $P < .001$), and WHR (+0.03, 95% CI =

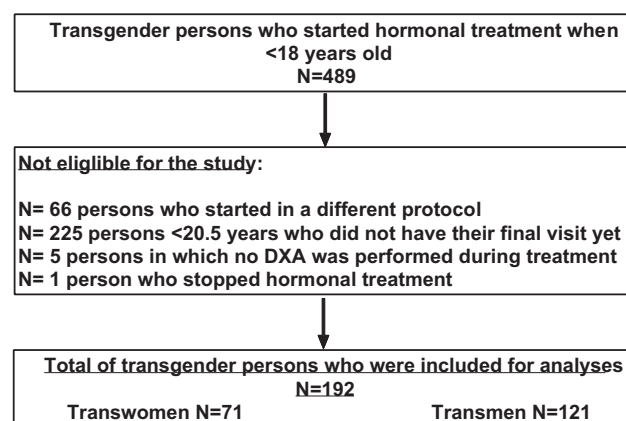


Figure 1. Flowchart of participant inclusion process. DXA = whole-body dual-energy x-ray absorptiometry.

Table 1. Characteristics of adolescents who started treatment with GnRHa and CHT at the VU University Medical Center from 1998 through 2014*

	Transwomen	Transmen
Adolescents	71	121
Age at start of GnRHa (y)	14.5 ± 1.8	15.3 ± 2.0
Age at start of CHT (y)	16.4 ± 1.1	16.9 ± 0.9
Ethnicity, %		
Caucasian	98	94
Asian	1	2
Black American	1	2
Occurrence of menarche, % [†]	—	84
BMI at start (kg/m ²)	19.8 (18.0–22.0)	20.6 (19.1–23.1)
Duration of GnRHa monotherapy (y)	2.1 (1.0–2.8)	1.0 (0.5–2.9)
Duration of GnRHa + CHT (y)	3.1 (2.5–3.6)	2.4 (2.0–3.1)
Duration of CHT monotherapy (y)	2.8 (1.6–3.4)	3.0 (1.9–3.4)
E2 level at start of GnRHa (pmol/L)	50 (20–79)	112 (70–202)
E2 level at start of CHT (pmol/L)	25 (20–31)	28 (23–36)
E2 level at 22 y of age (pmol/L)	121 (81–154)	70 (43–135)
T level at start of GnRHa (nmol/L)	10.0 (4.3–14.0)	1.0 (1.0–1.3)
T level at start of CHT (nmol/L)	1.0 (1.0–1.0)	1.0 (1.0–1.0)
T level at 22 y of age (nmol/L)	1.0 (0.8–1.0)	16.0 (8.8–37.0)
SDS at start vs ciswomen		
BMI	0.1 (–0.1 to 0.4)	0.4 (0.4–0.8)
Waist	0.6 (0.3–0.9)	0.6 (0.3–0.8)
Hip	0.0 (–0.2 to 0.2)	1.1 (0.1–0.5)
WHR	0.9 (0.6–0.9)	0.2 (0.0–0.3)
Total body fat	–0.9 (–1.0 to –0.7)	–0.1 (–0.2 to 0.1)
Lean body mass	2.5 (2.0–2.8)	1.3 (1.0–1.5)
SDS at start vs cismen		
BMI	0.4 (0.1–0.7)	0.7 (0.7–1.1)
Waist	0.1 (–0.2 to 0.4)	0.1 (–0.2 to 0.3)
Hip	0.3 (0.0–0.6)	0.7 (0.4–0.8)
WHR	–0.2 (–0.3 to 0.0)	–1.0 (–1.0 to –0.7)
Total body fat	1.6 (1.5–1.8)	2.0 (1.9–2.0)
Lean body mass	–0.7 (–0.9 to –0.6)	–1.1 (–1.1 to –1.0)

BMI = body mass index; CHT = cross-sex hormonal treatment; E2 = estradiol; GnRHa = gonadotropin-releasing hormone analogues; SDS = SD score; T = testosterone; WHR = waist-to-hip ratio.

*Data are presented as number, mean ± SD, median (interquartile range), or SDS (95% CI).

[†]Data were missing for 8% of transmen.

0.01–0.04, $P < .002$). Percentage of total body fat decreased (–3%, 95% CI = –4 to –2, $P < .001$), so percentage of LBM increased (+3%, 95% CI = 2–4, $P < .001$; [Table 2](#), [Figure 2](#)). Percentage of body fat decreased in the gynoid region (–5%, 95% CI = –6 to –3, $P < .001$) with no change in the android region (+1%, 95% CI = 0–3, $P = .18$; [Table 2](#)). Changes in body composition and body shape measurements were not different after adjustment for Tanner stage at start of treatment, or BMI at start of treatment.

Body Shape and Body Composition at 22 Years Compared With Peers

SDS of body shape and body composition in transgender persons for both ciswomen and cismen are presented in [Table 3](#).

Influence of Tanner Stage at Start on Body Shape and Body Composition at 22 Years

Transmen who started CHT in early or mid-puberty had a higher WHR than transmen who started treatment in late puberty. Transwomen tended to have a lower WHR starting in early or mid-puberty than those who started in late puberty. No differences in percentage of total body fat or percentage of total LBM were found across Tanner stages at start of treatment in transwomen and transmen ([Table 4](#)).

DISCUSSION

This study of 71 transwomen and 121 transmen shows that measurements of body shape and body composition change

Table 2. Measurements of body shape and body composition at the start of GnRHa, the start of CHT, and at 22 years of age in transwomen (n = 71) and transmen (n = 121)

	Start of GnRH	Start of CHT	22 y of age
Transwomen			
Body weight (kg)	58 (56–61)	66 (63–69)	76 (71–82)
BMI (kg/m ²)	20.2 (19.4–20.9)	21.3 (20.5–22.0)	23.2 (21.6–24.8)
Waist circumference (cm)*	71 (69–73)	75 (73–77)	79 (77–82)
Hip circumference (cm)*	89 (87–91)	95 (93–97)	106 (102–110)
WHR*	0.81 (0.79–0.82)	0.79 (0.78–0.80)	0.77 (0.75–0.79)
Body fat			
Total body (%)*	25 (23–26)	31 (29–32)	34 (32–36)
Android (%)*	23 (21–25)	28 (26–31)	32 (28–36)
Gynoid (%)*	29 (27–30)	36 (34–38)	40 (38–42)
Lean body mass			
Total body (%)*	75 (74–77)	69 (68–71)	66 (64–68)
Transmen			
Body weight (kg)	58 (56–61)	63 (60–65)	69 (66–71)
BMI (kg/m ²)	21.6 (20.9–22.3)	22.5 (21.7–23.2)	23.9 (23.0–24.7)
Waist circumference (cm)*	71 (69–72)	73 (71–74)	77 (75–79)
Hip circumference (cm)*	92 (90–93)	95 (93–97)	96 (94–99)
WHR*	0.77 (0.76–0.78)	0.76 (0.75–0.77)	0.80 (0.78–0.82)
Body fat			
Total body (%)*	30 (29–31)	33 (32–35)	27 (26–28)
Android (%)*	29 (27–30)	33 (32–35)	30 (28–32)
Gynoid (%)*	36 (35–37)	39 (38–40)	31 (30–33)
Lean body mass			
Total body (%)*	70 (69–71)	67 (66–68)	73 (72–74)

BMI = body mass index; CHT = cross-sex hormonal treatment; GnRHa = gonadotropin-releasing hormone analogues; WHR = waist-to-hip ratio.

*Percentages of missing data for anthropometrics were 11% in transwomen and 18% in transmen for start of GnRHa, 10% in transwomen and 13% in transmen for start of CHT, and 71% in transwomen and 76% for visit at 22 years of age. For measurements of body composition examined by whole-body dual-energy x-ray absorptiometry, percentages of missing data were 12% in transwomen and 11% in transmen for start of GnRHa, 36% in transwomen and 45% in transmen for start of CHT, and 64% in transwomen and 65% in transmen at 22 years of age.

toward the values of the affirmed sex during treatment with GnRHa and CHT. As a result of these changes, in young adult transwomen at 22 years of age, SDS for WHR, body fat, and LBM showed greater similarity to ciswomen than to cismen. In transmen at the same age, SDS for WHR, body fat, and LBM were between reference values for ciswomen and cismen. The achieved body fat and LBM at 22 years were similar across different Tanner stages at start of treatment in transwomen and transmen. However, in transmen, an earlier Tanner stage at start of treatment appeared to be associated with a closer resemblance of body shape to their affirmed sex at 22 years, and this tended to be the same in transwomen.

Compared with ciswomen and cismen in adolescence, transgender adolescents who are treated with GnRHa and CHT exhibit greater changes in body composition. A larger increase in percentage of body fat has been seen in transwomen compared with ciswomen within the same lifespan.¹⁰ Transmen exhibit decreased percentage of body fat, whereas percentage of body fat in cismen remains stable during puberty.¹⁰ This observation is likely explained by the fact that prepubertal girls already have

more body fat than prepubertal boys, and therefore transgender persons have a different body composition at the start of hormonal treatment than age-matched peers of the affirmed sex.¹³ Also, compared with adult transgender persons treated with CHT,^{14,15} larger changes in body shape and body composition are seen in transgender persons who start in adolescence. Moreover, transgender persons who start treatment in adolescence have at 22 years of age a body composition that approaches the affirmed sex more closely than transgender persons who start CHT in adulthood.¹⁵

The appropriate moment of starting gender reassignment continues to be a topic of debate.¹⁶ The findings of this study favor starting treatment in an early stage of puberty, because this appears to be associated with a closer resemblance of body shape to the affirmed sex at 22 years. This observation can be explained in part by the fact that in early puberty there is no distinct sex-specific body fat distribution, but also other factors can contribute to this difference. For instance, the period of gonadal suppression is generally considered a period of status quo. However, it can be postulated that a body of a transwomen

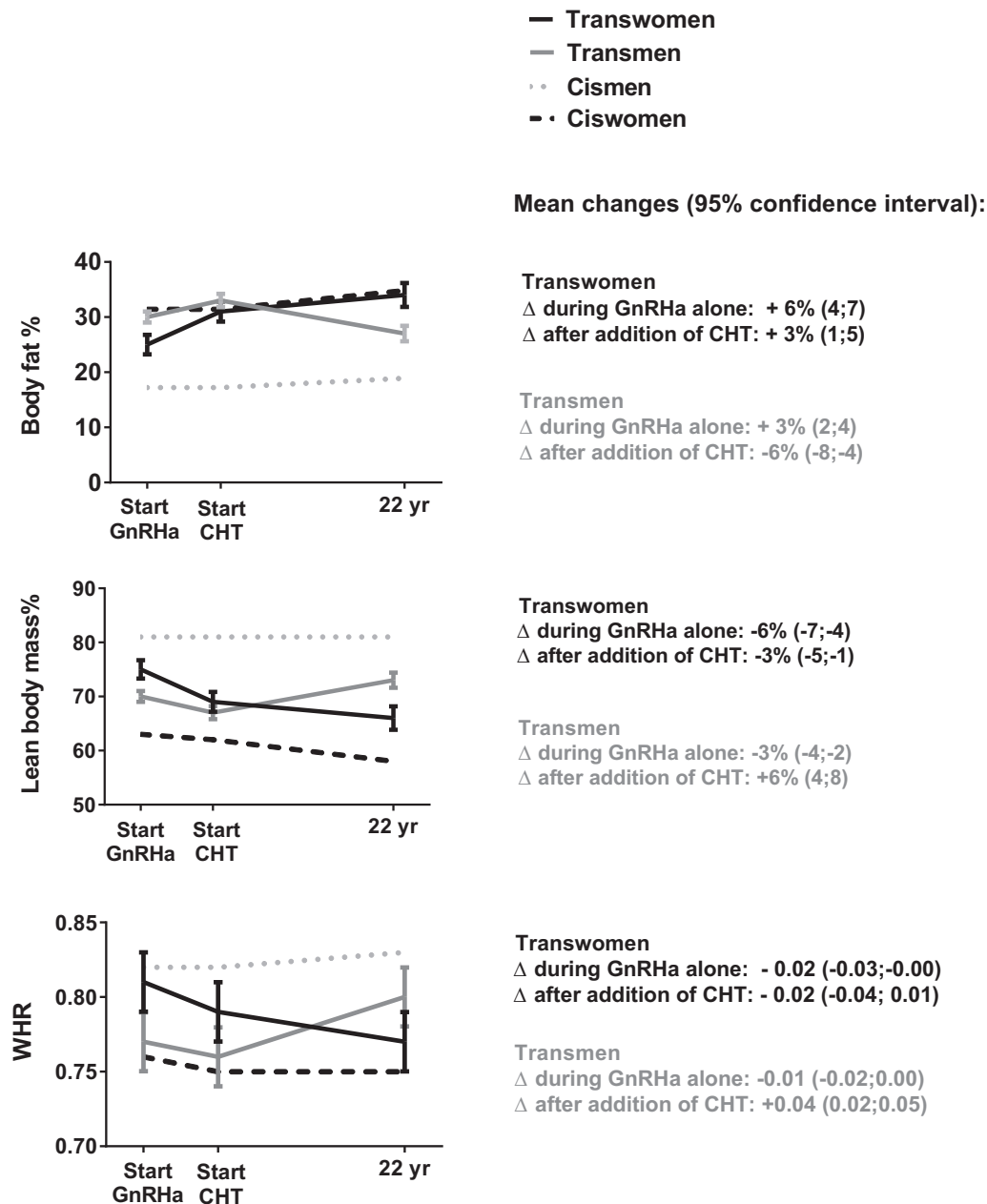


Figure 2. Changes (+95% CI) in percentage of total body fat, percentage of total body lean mass, and WHR in transwomen and transmen from the start of GnRHa to 22 years of age examined with mixed-model analyses using reference values for ciswomen and cismen. Reference data for ciswomen and cismen are shown for percentage of body fat,¹⁰ percentage of lean body mass,¹⁰ and WHR.⁴ Mean differences between changes during GnRHa alone and changes after the addition of CHT in transwomen were +3 (95% CI -0 to 5) for percentage of total body fat, -3 (95% CI -5 to 0) for percentage of lean body mass, and 0.00 (95% CI -0.02 to 0.03) for WHR and in transmen were +9 (95% CI 7-11) for percentage of total body fat, -9 (95% CI -11 to -7) for percentage of lean body mass, and -0.04 (95% CI -0.06 to -0.02) for WHR. CHT = cross-sex hormonal treatment; GnRHa = gonadotrophin-releasing hormone analogues; WHR = waist-to-hip ratio.

in a prolonged androgen-deprived state might respond differently to estrogens than a body that has had greater androgen exposure. Although programming properties of androgens are most pronounced prenatally, during puberty programming effects have been described in the brain.¹⁷ Therefore, other sex dimorphic organs such as body fat might also undergo

androgen-induced programming that results in a lesser response to estrogens. To achieve a better outcome (ie, greater resemblance to the affirmed sex), both this phase of gonadal suppression can be modulated as the phase of gonadal suppression combined with CHT. Although in this study the period of CHT did not differ much between pubertal groups, previous

Table 3. SDS of body shape and body composition in trans persons at 22 years of age for ciswomen and cismen

	Transwomen (n = 71)		Transmen (n = 121)	
	SDS for ciswomen	SDS for cismen	SDS for ciswomen	SDS for cismen
BMI	+0.4 (−0.2 to 1.0)	+0.4 (−0.2 to 1.0)	+0.5 (0.3–0.8)	+0.6 (0.3–0.8)
Waist circumference	+1.1 (0.8–1.6)	−0.1 (−0.4 to 0.3)	+0.8 (0.5–1.1)	−0.1 (−0.4 to 0.3)
Hip circumference	+1.4 (1.0–1.9)	+1.8 (1.3–2.4)	+0.2 (−0.2 to 0.5)	+0.3 (−0.2 to 0.7)
WHR	+0.3 (0.0–0.5)	−1.0 (−1.3 to −0.7)	+0.6 (0.4–0.8)	−0.5 (−0.8 to −0.3)
Total body fat	−0.4 (−0.8 to 0.0)	+2.1 (2.1–2.3)	−1.1 (−1.4 to −0.9)	+1.8 (1.6–1.9)
Android fat	−0.1 (−0.3 to 0.3)	+0.8 (0.6–1.1)	−0.3 (−0.4 to 0.0)	+0.7 (0.5–0.9)
Gynoid fat	−0.1 (−0.4 to 0.3)	+2.1 (1.8–2.3)	−1.7 (−1.8 to −1.5)	+0.9 (0.8–1.1)
Lean body mass	+1.3 (0.8–1.8)	−1.4 (−1.5 to −1.2)	+3.0 (2.8–3.3)	−0.9 (−0.9 to −0.8)

BMI = body mass index; SDSs = SD scores; WHR = waist-to-hip ratio.

Table 4. Measurements of body shape and body composition at 22 years of age shown per category of Tanner stage at start of treatment and differences between those categories

	Tanner stages at start of treatment			Difference between Tanner stages	
	Early puberty (n = 16)	Mid-puberty (n = 21)	Late puberty (n = 34)	ΔEarly vs mid-puberty	ΔEarly vs late puberty
Transwomen					
Age at start (y)	13.2 (12.8–13.5)	13.3 (13.1–14.0)	15.6 (14.3–17.2)		
Duration of GnRHa monotherapy (y)	2.7 (2.4–3.1)	2.7 (2.1–3.1)	1.0 (0.6–1.6)		
Duration of GnRHa + CHT (y)	3.0 (2.6–3.6)	3.1 (2.6–3.7)	3.1 (2.3–3.5)		
BMI (kg/m ²)	23.8 (21.5–26.2)	20.9 (18.7–23.1)	24.2 (22.8–25.6)	2.9 (−0.3 to 6)	−0.4 (−3.1 to 2.3)
SDS for ciswomen	0.6 (−0.2 to 1.5)	−0.4 (−1.2 to 0.4)	2.1 (0.3–1.3)		
WHR	0.75 (0.71–0.79)	0.75 (0.69–0.80)	0.78 (0.76–0.80)	0.00 (−0.07 to 0.07)	−0.03 (−0.08 to 0.02)
SDS for ciswomen	0.0 (−0.5 to 0.5)	0.0 (−0.8 to 0.6)	0.4 (0.1–0.6)		
Total body fat (%)	33 (29–38)	30 (25–35)	34 (32–36)	3 (−4 to 10)	−1 (−6 to 4)
SDS for ciswomen	−0.2 (−0.8 to 0.4)	−0.6 (−1.4 to 0.0)	−0.1 (−0.4 to 0.1)		
Lean body mass (%)	67 (62–72)	70 (65–75)	66 (64–68)	−3 (−10 to 4)	+1 (−4 to 6)
SDS for ciswomen	1.5 (0.3–2.8)	2.3 (1.0–3.5)	1.3 (0.8–1.8)		
Transmen	Early to mid-puberty (n = 11)		Late puberty (n = 110)	ΔEarly to mid vs late puberty	
Age at start (y)	12.3 ± 0.5		15.6 ± 1.7		
Duration of GnRHa monotherapy (y)	3.5 (3.4–3.9)		0.9 (0.5–2.0)		
Duration of GnRHa + CHT (y)	2.2 (2.0–2.5)		2.5 (1.9–3.2)		
BMI (kg/m ²)	22.5 (19.8–25.1)		23.9 (23.0–24.7)	−1.4 (−4.2 to 1.4)	
SDS for cismen	0.1 (−0.9 to 1.1)		0.7 (0.3–1.0)		
WHR	0.85 (0.80–0.90)		0.79 (0.77–0.80)	0.06 (0.02 to 0.12)	
SDS for ciswomen	0.3 (−0.5 to 1.2)		−0.7 (−1.0 to −0.5)		
Total body fat (%)	26 (21–31)		27 (26–29)	−1 (−6 to 4)	
SDS for cismen	1.7 (1.3–2.0)		1.8 (1.7–1.9)		
Lean body mass (%)	74 (69–79)		73 (72–74)	1 (−4 to 6)	
SDS for cismen	−0.8 (−1.1 to −0.4)		−0.9 (−0.9 to −0.8)		

BMI = body mass index; CHT = cross-sex hormonal treatment; GnRHa = gonadotropin-releasing hormone analogues; SDS = SD score; WHR = waist-to-hip ratio.

studies have shown that in adult transwomen¹⁸—but not in transmen¹⁹—a longer period of CHT establishes ongoing changes in body composition after the 1st year of treatment, which also could be applicable to body fat distribution and body shape. Therefore, one can postulate that starting CHT earlier than the current recommended age of 16²⁰ would improve outcomes in young adulthood.

We observed that transwomen at 22 years more closely resembled the affirmed sex regarding total body fat than transmen did. Body composition is the result of an interplay among multiple factors, including hormonal status, genetic predisposition, and lifestyle features such as diet and physical activity levels. A variance in these factors could explain this difference in treatment results between transwomen and transmen. Before the start of GnRHa, this difference was already present. Transmen had a percentage of total body fat very similar to that of ciswomen (SDS = -0.1), but transwomen had a percentage of body fat closer to that of ciswomen (SDS = -0.9) than to that of cismen (SDS = 1.6). The cause of the increased percentage of body fat and BMI in transwomen is unknown, but it can be postulated that psychological stress from gender dysphoria and an inactive lifestyle²¹ could have contributed. Alternatively, the genetic predisposition, independent of hormonal treatment, of transmen could have resulted in less comparable values to their affirmed sex at 22 years than that of transwomen.²² It can be debated whether the 46,XX karyotype²³ or other autosomal genes²⁴ are responsible.

As shown in Figure 2, the largest changes in transwomen were seen during GnRHa monotherapy (ie, during the suppression of testosterone). In transmen, the largest change was observed after the start of testosterone treatment. Therefore, it can be postulated that the suppression or addition of testosterone has a greater impact than the suppression or addition of estradiol. However, it is unclear whether the larger increase in body fat during GnRHa monotherapy in transwomen is due solely to the direct absence of testosterone action or the hypogonadal state itself. Indeed, also in transmen an increase in body fat was seen during GnRHa monotherapy. The larger change in body fat in transmen after the addition of testosterone could be due to the known lipolytic and anabolic effects of testosterone,²⁵ which in turn could be enhanced by practicing more sports or strength training in later adolescence.

This study is the 1st examining the effects of GnRHa and CHT on body shape and body composition in a large cohort of transgender adults who started treatment in their teens. A limitation of our study is the presence of missing data, which is seen more often in studies with a retrospective design when data are collected during regular patient care. However, our missing data were missing at random and analyses between persons with missing data and persons without missing data did not indicate selection bias. Another limitation of our study is that

(change in) body fat and LBM are dependent on other factors such as diet and physical activity and those were not systematically recorded.

This study provides insight into the effects of GnRHa and CHT on body shape and body composition during treatment in adolescence. Transgender adolescents starting treatment can be better informed on the extent to which their bodies will change and on what to expect of the results of such changes in young adulthood. Also, the finding that an early start of treatment in transmen results in a body shape that more closely resemble those of the affirmed sex than a late start supports early medical intervention.

The transgender persons included in this study were the first adolescents to be treated with the Dutch protocol.¹ Despite the favorable results on body shape and body composition, future research is warranted to improve current treatment protocols or devise alternative treatment protocols when GnRHa are not available.²⁶ Further, this study focused on body shape and body composition as important features of physical appearance, but it would also be of interest to determine the effects on body composition in the context of cardiovascular risk. For example, visceral fat seems to play an important role in the onset of insulin resistance and the metabolic syndrome,^{27,28} and thus the study of changes in subcutaneous and visceral fat depots and their relation to cardiometabolic outcome could be relevant in the future.

CONCLUSION

Transwomen who started treatment with GnRHa and CHT in adolescence showed an increase in body fat with a decrease in LBM and WHR. At 22 years of age, their body shape and body composition were more similar to those of ciswomen than to those of cismen. In transmen, an increase in LBM and WHR was seen with a decrease in body fat during therapy, and at 22 years they had values between reference values for ciswomen and cismen. The achieved body fat and LBM at 22 were similar across different Tanner stages at start of treatment in transwomen and transmen. However, in transmen, start of treatment in early or mid-puberty resulted in a body shape more similar to that of the affirmed sex, and this tended to be the same in transwomen.

Corresponding Author: D. T. Klink, MD, PhD, Department of Pediatrics, Division of Endocrinology, VU Medical Center, De Boelelaan 1117, 1081 HZ Amsterdam, The Netherlands. Tel: 0031-(0)-20442550; Fax: 0031-(0)-204445254; E-mail: d.klink@vumc.nl

Conflicts of Interest: The authors report no conflicts of interest.

Funding: None.

STATEMENT OF AUTHORSHIP

Category 1

(a) Conception and Design

M. Klaver; R. de Mutsert; D. T. Klink

(b) Acquisition of Data

M. Klaver; C. M. Wiepjes; J. Rotteveel; D. T. Klink

(c) Analysis and Interpretation of Data

M. Klaver; R. de Mutsert; J. W. R. Twisk; M. den Heijer;
D. T. Klink

Category 2

(a) Drafting the Article

M. Klaver; R. de Mutsert; D. T. Klink

(b) Revising It for Intellectual Content

M. Klaver; R. de Mutsert; C. M. Wiepjes; J. W. R. Twisk;
M. den Heijer; J. Rotteveel; D. T. Klink

Category 3

(a) Final Approval of the Completed Article

M. Klaver; R. de Mutsert; C. M. Wiepjes; J. W. R. Twisk;
M. den Heijer; J. Rotteveel; D. T. Klink

REFERENCES

- Delemarre-van de Waal HA, Cohen-Kettenis PT. Clinical management of gender identity disorder in adolescents: a protocol on psychological and paediatric endocrinology aspects. *Eur J Endocrinol* 2006;155:131-137.
- Wells JCK. Sexual dimorphism of body composition. *Best Pract Res Clin Endocrinol Metab* 2007;21:415-430.
- Taylor RW, Grant AM, Williams SM, et al. Sex differences in regional body fat distribution from pre- to postpuberty. *Obesity* 2010;18:1410-1416.
- Fredriks AM, van Buuren S, Fekkes M, et al. Are age references for waist circumference, hip circumference and waist-hip ratio in Dutch children useful in clinical practice? *Eur J Pediatr* 2005;164:216-222.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed, text rev. Washington, DC: American Psychiatric Association; 2000.
- Twisk JWR, de Boer M, de Vente W, et al. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. *J Clin Epidemiol* 2013; 66:1022-1028.
- Tanner JM, Whitehouse RH. Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. *Arch Dis Child* 1976;51:170-179.
- Prader A. Testicular size: assessment and clinical importance. *Triangle* 1966;7:240-243.
- Fredriks AM, van Buuren S, Wit JM, et al. Body index measurements in 1996–7 compared with 1980. *Arch Dis Child* 2000;82:107-112.
- van der Sluis IM, de Ridder MAJ, Boot AM, et al. Reference data for bone density and body composition measured with dual energy x ray absorptiometry in white children and young adults. *Arch Dis Child* 2002;87:341-347.
- Stults-Kolehmainen MA, Stanforth PR, Bartholomew JB. Fat in android, trunk, and peripheral regions varies by ethnicity and race in college aged women. *Obesity* 2012; 20:660-665.
- Stults-Kolehmainen MA, Stanforth PR, Bartholomew JB, et al. DXA estimates of fat in abdominal, trunk and hip regions varies by ethnicity in men. *Nutr Diabetes* 2013;3:1-6.
- Taylor RW, Gold E, Manning P, et al. Gender differences in body fat content are present well before puberty. *Int J Obes* 1997; 21:1082-1084.
- Klaver M, Dekker MJHJ, de Mutsert R, et al. Cross-sex hormone therapy in transgender persons affects total body weight, body fat and lean body mass: a meta-analysis. *Andrologia* <https://doi.org/10.1111/and.12660>. E-pub ahead of print.
- Klaver M, Dekker MJHJ, Schreiner T, et al. Cross-sex hormone therapy and the effects on fat distribution in transgender persons. Presented at: 24th WPATH Symposium; Amsterdam, The Netherlands; 2016.
- de Vries ALC, Klink D, Cohen-Kettenis PT. What the primary care pediatrician needs to know about gender incongruence and gender dysphoria in children and adolescents. *Pediatr Clin North Am* 2016;63:1121-1135.
- Schulz KM, Sisk CL. The organizing actions of adolescent gonadal steroid hormones on brain and behavioral development. *Neurosci Biobehav Rev* 2016;70:148-158.
- Mueller A, Zollver H, Kronawitter D, et al. Body composition and bone mineral density in male-to-female transsexuals during cross-sex hormone therapy using gonadotrophin-releasing hormone agonist. *Clin Endocrinol Diabetes* 2011; 119:95-100.
- Mueller A, Haeberle L, Zollver H, et al. Effects of intramuscular testosterone undecanoate on body composition and bone mineral density in female-to-male transsexuals. *J Sex Med* 2010;7:3190-3198.
- Hembree WC, Cohen-Kettenis P, Delemarre-van de Waal HA, et al. Endocrine treatment of transsexual persons: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2009;94:3132-3154.
- Jones BA, Arcelus J, Bouman WP, et al. Barriers and facilitators of physical activity and sport participation among young transgender adults who are medically transitioning. *Int J Transgend* 2017;18:227-238.
- Samaras K, Spector TD, Nguyen TV, et al. Independent genetic factors determine the amount and distribution of fat in women after the menopause. *J Clin Endocrinol Metab* 1997;82:781-785.
- Arnold AP. The organizational-activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Horm Behav* 2009; 55:570-578.

24. Chagnon YC, Rice T, Pérusse L, et al. Genomic scan for genes affecting body composition before and after training in Caucasians from HERITAGE. *J Appl Physiol* 2001; 90:1777-1787.
25. Geer EB, Shen W. Gender differences in insulin resistance, body composition, and energy balance. *Gend Med* 2009; 6:60-75.
26. Tack LJW, Heyse R, Craen M, et al. Consecutive cyproterone acetate and estradiol treatment in late-pubertal transgender female adolescents. *J Sex Med* 2017;14:747-757.
27. Després JP, Lemieux I. Abdominal obesity and metabolic syndrome. *Nature* 2006;444:881-887.
28. Tchernof A, Després JP. Pathophysiology of human visceral obesity: an update. *Physiol Rev* 2013;93:359-404.

Proandrogenic and Antiandrogenic Progestins in Transgender Youth: Differential Effects on Body Composition and Bone Metabolism

Lloyd J. W. Tack,¹ Margarita Craen,¹ Bruno Lapauw,^{2,3} Stefan Goemaere,^{2,3} Kaatje Toye,^{2,3} Jean-Marc Kaufman,² Sara Vandewalle,² Guy T'Sjoen,² Hans-Georg Zmierzczak,^{2,3} and Martine Cools¹

¹Division of Pediatric Endocrinology, Department of Pediatrics, Ghent University Hospital, Ghent University, 9000 Ghent, Belgium; ²Department of Endocrinology, Ghent University Hospital, Ghent University, 9000 Ghent, Belgium; and ³Unit for Osteoporosis and Metabolic Bone Diseases, Department of Internal Medicine, Ghent University Hospital, 9000 Ghent, Belgium

Context: Progestins can be used to attenuate endogenous hormonal effects in late-pubertal transgender (trans) adolescents (Tanner stage B4/5 and G4/5). Currently, no data are available on the effects of progestins on the development of bone mass or body composition in trans youth.

Objective: To study prospectively the evolution of body composition and bone mass in late-pubertal trans adolescents using the proandrogenic or antiandrogenic progestins lynestrenol (L) and cyproterone acetate (CA), respectively.

Design and Outcome Measurements: Forty-four trans boys (Tanner B4/5) and 21 trans girls (Tanner G4/5) were treated with L or CA for 11.6 (4 to 40) and 10.6 (5 to 31) months, respectively. Anthropometry, grip strength, body composition, and bone mass, size, and density were determined by dual-energy X-ray absorptiometry and peripheral quantitative computed tomography before the start of progestin and before addition of cross-sex hormones.

Results: Using L, lean mass [+3.2 kg (8.6%)] and grip strength [+3 kg (10.6%)] significantly increased, which coincided with a more masculine body shape in trans boys. Trans girls showed loss of lean mass [−2.2 kg (4.7%)], gain of fat mass [+1.5 kg (9.4%)], and decreased grip strength Z scores. CA limited normal bone expansion and impeded pubertal bone mass accrual, mostly at the lumbar spine [Z score: −0.765 to −1.145 ($P = 0.002$)]. L did not affect physiological bone development.

Conclusion: Proandrogenic and antiandrogenic progestins induce body composition changes in line with the desired appearance within 1 year of treatment. Bone health, especially at the lumbar spine, is of concern in trans girls, as bone mass accrual is severely affected by androgen suppressive therapy. (*J Clin Endocrinol Metab* 103: 2147–2156, 2018)

Worldwide, the number of children and adolescents presenting with gender incongruence (GI) is increasing (1). In children with persistent GI,

gonadotropin-releasing hormone analogs (GnRHa), after multidisciplinary evaluation and a variable period of psychological support, are commonly used to

ISSN Print 0021-972X ISSN Online 1945-7197

Printed in USA

Copyright © 2018 Endocrine Society

Received 20 October 2017. Accepted 11 April 2018.

First Published Online 16 April 2018

Abbreviations: aBMD, areal bone mineral density; BMC, bone mineral content; BMI, body mass index; CA, cyproterone acetate; CSH, cross-sex hormones; DXA, dual-energy X-ray absorptiometry; GI, gender incongruence; GnRHa, gonadotropin-releasing hormone analogs; L, lynestrenol; NHANES, National Health and Nutrition Examination Survey; PINP, procollagen type I N-terminal propeptide; PTH, parathyroid hormone; pQCT, peripheral quantitative computed tomography; s-CTX, serum C-terminal telopeptide; SHBG, sex hormone-binding globulin; SSIP, polar strength strain index; trans, transgender; UCa/Cr, urinary calcium/creatinine ratio; vBMD, volumetric bone mineral density; WHR, waist/hip ratio.

arrest pubertal development. GnRHa suppress gonadotropin secretion and thus the gonadal production of sex hormones and the development of secondary sex characteristics of the gender recorded at birth (2). GnRHa have been reported to increase fat mass and reduce lean mass in both transgender (trans) boys and trans girls (3). In addition, concerns have been raised about the negative impact of this treatment on bone health, particularly bone mass accrual in trans girls, which may not be fully compensated by subsequent addition of cross-sex hormones (CSH) (4, 5).

If secondary sex characteristics have already developed at the time GI is diagnosed, the focus of treatment shifts toward reducing the burden of isosexual pubertal signs, such as menstruation in trans boys or erections and facial and body hair growth in trans girls, while awaiting the start of CSH treatment—usually around the age of 16 years, although strict application of this age criterion is under debate. Apart from GnRHa treatment, menstruation in girls and erections and hair growth in boys can also successfully be suppressed with proandrogenic and antiandrogenic progestins such as lynestrenol (L) or cyproterone acetate (CA), respectively, provided that these drugs are available (6, 7). The low cost of progestins is an advantage over the expensive GnRHa, especially in circumstances in which these drugs are not reimbursed.

L is a proandrogenic progestin that increases the endogenous serum testosterone/estradiol ratio, thereby interfering with the menstrual cycle and inducing amenorrhea, especially at higher doses (10 mg/d) (6). The antiandrogenic progestin CA is a competitive androgen receptor blocker resulting in decreased androgenic effects (7–9). Antiandrogens such as CA have previously been shown to decrease lean body mass and increase total body fat in adult trans women (10, 11), and in some adult transgender teams, proandrogenic and antiandrogenic progestins are used prior to CSH or in association with CSH as an adjuvant therapy (10, 12). It has not been investigated whether the proandrogenic and antiandrogenic effects of L and CA affect anthropometry and body composition in trans adolescents and/or have less deleterious effects on bone mass accrual as compared with GnRHa treatment. If these progestins indeed induce bodily changes toward the desired appearance in addition to suppressing endogenous pubertal signs in late pubertal adolescents, they may, in line with adult treatment regimens, represent a first step toward CSH treatment in this specific population.

The aim of this study is to prospectively assess anthropometry, body composition, and bone mass changes in late-pubertal trans boys and trans girls who, respectively, received L or CA to attenuate endogenous

pubertal hormone effects while awaiting CSH therapy. This will contribute to determining the place of GnRHa and progestins, respectively, in the pharmacological treatment of trans youth and to improving our knowledge on the long-term effects of these interventions, as has been suggested recently (2).

Subjects and Methods

Participants and treatments

Sixty-five late-pubertal (Tanner stage B4/G4) trans adolescents (44 trans boys and 21 trans girls) who used progestins between March 2011 and January 2017 were included in the study. The cohort of participants in this study partially overlaps with cohorts of our previous reports on progestin-induced biochemical and hormonal changes in trans adolescents (6, 7). According to our local protocol, CSH can be added to this treatment from the age of 16 years onward and after a period of progestin use of ~6 months. In all subjects presenting before or in early puberty (Tanner stages B2 to 3/G2 to 3), GnRHa instead of progestins are used to arrest further pubertal development. Biochemical parameters and data on muscle strength, body composition, and bone mass were collected at the start of L/CA (baseline) and before addition of CSH.

The diagnosis of GI was made by a child psychologist and child psychiatrist and based on the diagnostic criteria as described in the *Diagnostic and Statistical Manual of Mental Disorders* 4th edition, and, later, the 5th edition, and the *International Classification of Diseases* 10th edition (13). The decision to start medical treatment was made by the multidisciplinary child gender team of Ghent University Hospital. A daily dose of 5 to 10 mg L (Orgametril®; MSD Belgium, Brussels, Belgium) or 50 mg CA (Androcur®; Bayer, Leverkusen, Germany) was given orally in trans boys and trans girls, respectively. In all individuals, vitamin D supplementation [cholecalciferol 25,000 IU/mo (D-CURE®; SMB Laboratories, Brussels, Belgium)] was added to the treatment, and all were advised to pursue a daily dietary calcium intake of at least 1500 mg.

Methods

Investigations were performed twice (*i.e.*, before initiation of L or CA and prior to addition of CSH) and consisted of the following: anthropometric measurements, a fasting venous blood sample (between 0800 h and 0900 h), a spot urine sample, dual-energy X-ray absorptiometry (DXA), peripheral quantitative computed tomography (pQCT) scans, and grip-strength measurements.

Approval of the study by the ethics committee of Ghent University Hospital was obtained (B670201525328).

Anthropometrics

Body weight was measured to the nearest 0.1 kg in light, indoor clothing without shoes. A wall-mounted Harpenden stadiometer (Holtain, Ltd, Crymch, United Kingdom) was used to measure height. Waist circumference was measured between the lower rib and the iliac crest at the end of expiration; hip circumference was measured at the widest gluteal circumference. Both measurements were performed twice. The mean of both measurements was used for analysis. SD scores of height,

weight, and body mass index (BMI) were calculated based on published data obtained in Flemish peers of the same gender recorded at birth (14).

Serum analyses

Sex hormone-binding globulin (SHBG), total testosterone, free testosterone, total estradiol, total testosterone/estradiol ratio, follicle-stimulating hormone, luteinizing hormone, serum 25-OH vitamin D, parathyroid hormone (PTH), serum C-terminal telopeptide (s-CTX), and procollagen type I N-terminal propeptide (PINP) were measured. Electrochemiluminescence assays (Roche Diagnostics E170 Modular; Roche Diagnostics, Indianapolis, IN) were used for all parameters, except total testosterone, for which liquid chromatography–tandem mass spectrometry was used. Free testosterone was calculated as described by Vermeulen *et al.* (15). The detection limit for estradiol was 25 ng/L; in case of values below this limit, the limit itself was input for statistical analysis. Urinary calcium/creatinine ratio (UCa/Cr) was calculated from the calcium and creatinine levels as measured on a urine spot sample.

Grip strength

An adjustable handheld standard grip device (JAMAR hand dynamometer; Sammons and Preston, Bolingbrook, IL) was used to determine grip strength on both the dominant and nondominant hand. Grip strength was measured three times on both hands. The highest value was assumed to represent the maximum strength and was input for analyses. Z scores were calculated based on data published by Ortega *et al.* (16).

DXA scans

A Hologic Discovery device (Software Version 11.2.1; Hologic, Inc., Bedford, MA) was used to assess body composition of the total body minus head (total, lean, and fat mass), bone mineral content (BMC), areal bone mineral density (aBMD) and projected bone area of the spine, left proximal femur (femoral neck and total hip), and whole body. Z scores of DXA scan parameters were based on data from the National Health and Nutrition Examination Survey (NHANES) and Bone Mineral Density in Childhood Study (17, 18). Z scores of aBMD were calculated by the manufacturer's software based on age-matched adolescents of the same gender recorded at birth; Z scores of percentage body fat and aBMD of the experienced gender were calculated manually based on the data from the NHANES study, kindly provided by Hologic, Inc.

pQCT scans

A pQCT device (XCT-2000; Stratec Medizintechnik, Pforzheim, Germany) was used to evaluate both the trabecular and cortical volumetric bone mineral density (vBMD) and bone geometry at the nondominant radius, and the left tibia measured at 4% and 66% and 4% and 38%, respectively, of bone length from distal. Polar strength strain index (SSI_p) was calculated based on geometrical and material properties at the nondominant radius at 66%. Muscle and fat area were measured at the midshaft radius of the nondominant arm and left tibia (at the 66% and 38% sites, respectively). Z scores of the radius were calculated by the manufacturer's software. Z scores of the tibia were calculated manually using the formulas published by Roggen *et al.* (19) based on results of the same pQCT device and enabled the calculation of Z scores of

periosteal and endosteal circumference (which are not available at the radius). All Z scores were based on age-matched adolescents of the same gender recorded at birth.

Manual calculation of Z scores was performed using the following formulae:

$$Z = \frac{M\left(\left(\frac{X}{M}\right)^L - 1\right)}{L\sigma} \text{ or } Z = \frac{X - M}{\sigma},$$

in which X is the physical measurement, L the age-matched skewness value, M the median age-matched reference value, and σ the age-matched population SD.

Statistical analyses

Statistical analysis was performed using an IBM SPSS software package (version 23). A P value of <0.05 was considered significant. Mean values and SD or medians and interquartile ranges are given, with Gaussian and non-Gaussian distributions, respectively (based on a Shapiro-Wilk test). A paired Student *t* test or Wilcoxon signed-ranks test was used after testing for normality with a Shapiro-Wilk test, to compare data prior to initiation of progestins and before initiation of CSH. Four of the 44 trans boys used an oral contraceptive prior to starting with L. Data for these adolescents were excluded from the hormonal analyses. Because unpaired Student *t* tests/Mann-Whitney *U* test (depending on the results of a Shapiro-Wilk test) revealed similar body composition parameters and similar changes in bone parameters, those data were included in all other analyses.

Results

Sixty-five adolescents (44 trans boys and 21 trans girls) were included in the study. At the start of progestin treatment, mean chronological and bone age were 16.2 ± 1.05 and 16.4 ± 1.08 years in trans boys and 16.3 ± 1.21 and 17.1 ± 1.28 years in trans girls, respectively (Table 1). The mean time interval between both examinations was 11.64 (4 to 40) months in trans boys and 10.57 (5 to 31) months in trans girls, respectively.

Serum analyses

Mean total testosterone and estradiol levels decreased with progestin treatment in both groups (Table 2). However, the total testosterone-to-estradiol ratio decreased significantly in trans girls ($P = 0.008$) but not in trans boys ($P = 0.068$). Because of the decreased SHBG levels, free testosterone levels increased in trans boys. In trans girls, SHBG remained stable,

Table 1. Descriptive Statistics

	Trans Boys	Trans Girls
Number of participants	44	21
Age at start of progestin, y	16.2 ± 1.05	16.3 ± 1.21
Bone age at start of progestin, y	16.4 ± 1.08	17.1 ± 1.28

Table 2. Summary of Serum Markers

Serum markers	Trans Boys			Trans Girls		
	Before L	After L	P (L)	Before CA	After CA	P (CA)
Total testosterone, ng/dL	33.00 (17.39)	23.85 (19.86)	0.039 ^a	438.21 ± 211.07	253.35 ± 172.96	<0.001 ^a
Free testosterone, ng/dL	0.41 (0.22)	0.62 (0.33)	0.024 ^a	9.64 (± 4.40)	5.69 (± 3.09)	0.010 ^a
Total estradiol, ng/L	53.20 (58.43)	25.00 (13.75)	<0.001 ^a	25.8 (11.0)	25.0 (0)	0.002 ^a
T/E ratio	0.518 (0.800)	0.824 (0.710)	0.068	15.593 ± 5.751	11.024 ± 6.557	0.008 ^a
SHBG, nmol/L	54.7 (54.38)	22.10 (14.15)	<0.001 ^a	28.44 ± 9.54	26.38 ± 9.23	0.366
LH, U/L	7.60 (7.73)	2.90 (3.9)	<0.001 ^a	3.77 ± 1.33	5.33 ± 2.30	0.007 ^a
FSH, U/L	5.96 ± 2.06	4.49 ± 2.66	0.009 ^a	3.1 (2.60)	2.1 (4.85)	0.614
Vitamin D, ng/mL	17.72 ± 8.96	23.20 ± 7.89	0.001 ^a	17.34 ± 6.66	24.82 ± 8.37	0.008 ^a
PTH, ng/L	44.2 (11.6)	30.95 (16.05)	0.001 ^a	40.15 (24.10)	26.9 (9.7)	0.113
PINP, µg/L	140.10 (82.12)	127.05 (74.65)	0.018 ^a	232.0 (377.7)	124.10 (137.9)	0.005 ^a
s-CTX, ng/mL	0.784 ± 0.266	0.799 ± 0.273	0.747	1.211 ± 0.466	1.004 ± 0.318	0.042 ^a
UCa/Cr, mg/mg	0.095 ± 0.056	0.051 ± 0.029	<0.001 ^a	0.054 ± 0.045	0.080 ± 0.067	0.035 ^a

Data are mean ± SD or median (interquartile range).

Abbreviations: After L or CA, mean/median levels before start of CSH; Before L or CA, mean/median values before start of progestin; FSH, follicle-stimulating hormone; LH, luteinizing hormone; T/E ratio, testosterone/estradiol ratio; Vitamin D, 25-OH vitamin D.

^aP < 0.05.

resulting in decreased free testosterone levels ($P = 0.010$). Gonadotropin levels decreased only in trans boys. Serum levels of *PINP*, a marker for bone formation, decreased by 9.3% in trans boys and 46.5% in trans girls. In the latter group, s-CTX levels, a marker for bone resorption, also decreased by 17.1%. Mean vitamin D levels before initiation of treatment were in the range of vitamin D insufficiency (*i.e.*, <20 ng/mL): 17.72 ng/mL in trans boys and 17.34 ng/mL in trans girls; mean levels increased in both groups following supplementation. PTH levels decreased in both groups; however, this was only significant in trans boys *versus* trans girls ($P = 0.001$ and 0.113, respectively). UCa/Cr

showed a decrease in trans boys ($P < 0.001$) and an increase in trans girls ($P = 0.035$).

Anthropometrics

Trans boys grew on average 0.8 cm and gained 2.65 kg during the study period; weight and BMI gain were more pronounced than in age-matched female peers ($P = 0.001$) (Table 3). Both waist and hip circumference increased ($P < 0.001$ and 0.002) as well as waist/hip ratio (WHR) ($P = 0.041$).

Trans girls treated with CA grew on average 0.7 cm and gained 2.33 kg; weight and BMI changes were similar to age-matched male peers. Their waist and hip

Table 3. Anthropometrics

	Trans Boys			Trans Girls		
	Before L	After L	P (L)	Before CA	After CA	P (CA)
Reference height, ^a cm	165.9 ± 6.0	166.2 ± 6.0	—	176.5 ± 7.2	178.4 ± 7.0	—
Height, cm	165.2 ± 6.5	166.0 ± 6.5	<0.001 ^b	173.7 ± 7.1	174.4 ± 6.9	0.016 ^b
SD-score height	−0.039 ± 0.988	−0.025 ± 0.980	0.560	−0.177 ± 0.974	−0.441 ± 1.019	0.004 ^b
Reference weight, ^a kg	57.5 ± 6.7	58.6 ± 6.7	—	63.7 ± 8.0	66.3 ± 7.9	—
Weight, kg	58.50 (20.00)	60.50 (19.50)	0.001 ^b	66.24 ± 11.20	68.57 ± 13.18	0.167
SD-score weight	−0.041 ± 1.379	0.203 ± 1.265	0.001 ^b	0.413 ± 1.167	0.194 ± 1.255	0.118
Reference BMI, ^a kg/m ²	20.9 ± 2.2	21.1 ± 2.1	—	20.3 ± 1.9	20.7 ± 1.9	—
BMI, kg/m ²	21.50 (5.80)	22.35 (5.37)	<0.001 ^b	20.25 (4.45)	21.52 (2.84)	0.164
SD-score BMI	0.020 ± 1.400	0.306 ± 1.222	0.001 ^b	0.625 ± 1.025	0.506 ± 1.025	0.406
Waist circumference, cm	72.00 (13.81)	75.25 (13.00)	<0.001 ^b	75.00 (12.25)	76.15 (11.63)	0.586
Hip circumference, cm	96.23 ± 11.92	99.24 ± 9.77	0.002 ^b	91.53 ± 7.40	92.88 ± 9.17	0.462
WHR	0.749 (0.070)	0.781 (0.090)	0.041 ^b	0.810 (0.080)	0.828 (0.090)	0.306

Data are mean ± SD or median (interquartile range). Z scores are mean of Z scores of individual participants; age ranges: before L (11.9–18.4), after L (16.0–21.7), before CA (13.5–18.4), and after CA (16.0–19.1), based on data published by Roelants *et al.* (14).

Abbreviations: After L/CA, mean/median levels before start of CSH; Before L/CA, mean/median values before start of progestin.

^aThese rows represent reference values of age-matched peers of the same gender recorded at birth.

^bP < 0.05.

circumference only showed slight increases, resulting in a stable WHR.

Body composition and grip strength

In trans boys, lean mass (Fig. 1C), opposite to fat mass (Fig. 1D), significantly increased, leading to an increased total body mass (Fig. 1B) and decreased body fat percentage, significantly differing from their age-matched female peers, as calculated from population-based Z scores of the percentage of body fat ($P = 0.010$) (Supplemental Table 1). Muscle area increased at the nondominant forearm and left lower leg (radius, 66% and tibia, 38%). This increase at the forearm was higher than what was observed in age-matched peers of the same gender recorded at birth, based on Z scores ($P < 0.001$). There were no changes in forearm and lower leg fat area (Supplemental Table 1). An increase in grip strength was noted in the dominant and nondominant hand. This

increase was higher than in age-matched peers of the same gender recorded at birth ($P < 0.001$) (Fig. 1A; Supplemental Table 1).

Trans girls treated with CA showed a significant increase in fat mass (Fig. 1D) and decrease in lean mass (Fig. 1C), resulting in an increased body fat percentage ($P < 0.001$), without changes in total mass ($P = 0.167$) (Fig. 1B). This increase was higher than observed in their age-matched male peers, based on Z scores ($P = 0.001$) (Supplemental Table 1). At the nondominant forearm and left lower leg (radius, 66% and tibia, 38%), muscle area decreased, whereas fat area markedly increased. Based on Z scores, the changes of muscle and fat area at the forearm are different from their age-matched peers of the same gender recorded at birth (Supplemental Table 1). No significant changes in grip strength were observed in trans girls during the study period (Fig. 1A), resulting in decreased Z scores compared with age-matched

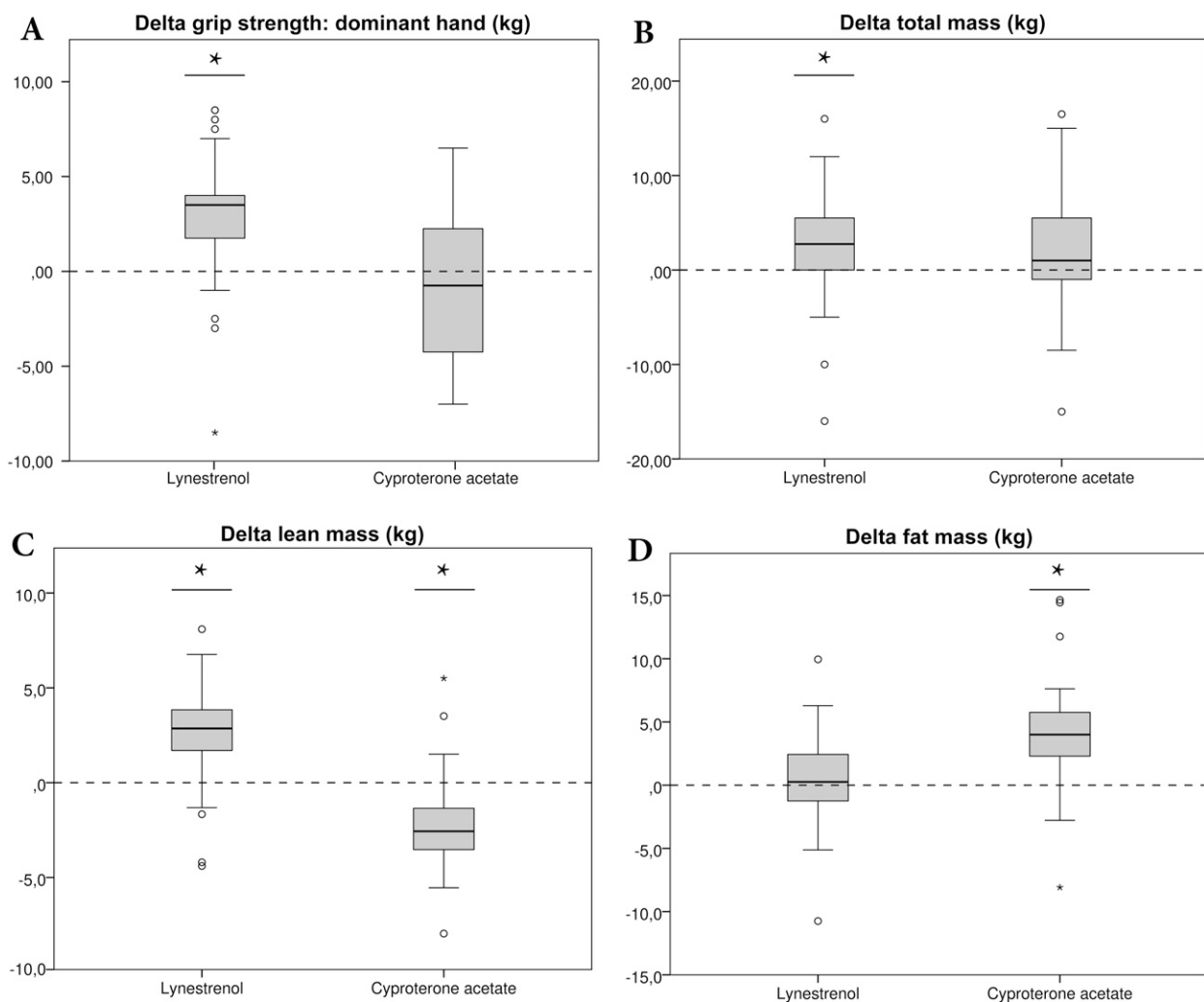


Figure 1. Changes in grip strength and body composition. Box and whisker plots of (A) delta grip strength of the dominant hand, (B) delta total mass, (C) delta lean mass, and (D) delta fat mass. Boxes indicate the 25th and 75th percentiles; horizontal lines within the shaded boxes indicate the median. Whiskers indicate the minimum and maximum; mild and extreme outliers (circles and stars, respectively) are defined as 1.5 and 3 times the interquartile range below or above the 25th and 75th percentile, respectively. * $P < 0.05$.

peers of the same gender recorded at birth ($P = 0.004$ and 0.056) (Supplemental Table 1).

Bone changes

In trans boys, aBMD of the total hip, femoral neck, lumbar spine, and whole body before the start of L were slightly lower compared with population-based references for the same age and gender recorded at birth (Z scores: -0.162 , -0.271 , -0.320 , and -0.783 , respectively) (Table 4).

At the total hip, an increase of BMC ($P = 0.004$) with a stable bone area was seen. This led to an increase in absolute values and Z scores of aBMD (both $P < 0.001$) (Table 4). At the femoral neck, Z scores of aBMD did not change significantly ($P = 0.057$) (Fig. 2C). At the lumbar spine, BMC and aBMD increased, whereas Z scores remained stable, suggesting an acquisition similar to age-matched female peers (Fig. 2B; Table 4).

Total body BMC, bone area, and aBMD increased ($P < 0.001$), whereas Z scores remained stable ($P = 0.720$) (Fig. 2A). Compared with age-matched peers of the experienced (male) gender, Z scores of aBMD evolved as follows: -0.95 to -1.04 ($P < 0.001$) for femoral neck, 0.04 to -0.24 ($P = 0.017$) for lumbar spine, and -1.01 to -1.29 ($P < 0.001$) for whole body.

Based on the pQCT scans, both trabecular and cortical bone parameters (area, BMC, and vBMD) at the radius and the tibia increased during the study period, to the

same extent as in age-matched control girls. SSIP Z scores at the midshaft radius did not change throughout the study period. All pQCT data are represented in Supplemental Table 2.

In trans girls, already at the start of CA, aBMD was lower as compared with their peers of the same gender recorded at birth (Z scores: -0.805 , -0.791 , -0.765 , and -1.381 for total hip, femoral neck, lumbar spine, and whole body, respectively) (Table 4). During the study period, DXA scans of the total hip and femoral neck showed an increase in bone area and a decrease in BMC, leading to a decreased aBMD, which was statistically significant only at the total hip ($P = 0.016$). Compared with their age-matched male peers, Z scores of the total hip and femoral neck decreased ($P = 0.006$ and 0.013 , respectively) (Fig. 2C; Table 4).

At the lumbar spine, bone area and BMC both increased, resulting in a stable aBMD but decreased aBMD Z scores ($P = 0.002$) (Fig. 2B; Table 4). At the whole-body level, increases in bone area and BMC were seen. However, aBMD did not change, resulting in decreased Z scores ($P = 0.008$) (Fig. 2A).

Compared with age-matched peers of the experienced (female) gender, aBMD Z scores further decreased from -0.45 to -0.67 ($P = 0.002$) for the femoral neck and showed negative trends from -1.11 to -1.29 ($P = 0.063$) for the lumbar spine and -0.92 to -1.01 ($P = 0.147$) for the whole body.

Table 4. Summary of DXA Measurements

	Trans Boys			Trans Girls		
	Before L	After L	P (L)	Before CA	After CA	P (CA)
Hip						
Total hip area, cm ²	33.826 ± 3.870	33.789 ± 3.948	0.774	38.742 ± 3.157	39.038 ± 2.765	0.341
BMC, g	29.356 (6.840)	30.060 (6.870)	0.004 ^a	35.941 ± 6.202	35.614 ± 5.450	0.383
Reference aBMD, ^b g/cm ²	0.957 ± 0.109	0.970 ± 0.108	—	1.029 ± 0.137	1.068 ± 0.143	—
aBMD, g/cm ²	0.926 ± 0.102	0.942 ± 0.104	<0.001 ^a	0.914 ± 0.125	0.897 ± 0.130	0.016 ^a
Z scores: BMD	-0.162 ± 0.888	-0.024 ± 0.889	<0.001 ^a	-0.805 ± 0.987	-1.010 ± 0.983	0.006 ^a
Femoral neck						
BMC, g	4.074 ± 0.675	4.163 ± 0.702	0.021 ^a	4.527 ± 0.784	4.518 ± 0.781	0.845
Reference aBMD, ^b g/cm ²	0.874 ± 0.106	0.883 ± 0.106	—	0.945 ± 0.131	0.975 ± 0.138	—
aBMD, g/cm ²	0.815 ± 0.103	0.826 ± 0.104	0.026 ^a	0.823 ± 0.128	0.811 ± 0.126	0.059
Z scores: BMD	-0.271 ± 0.906	-0.191 ± 0.933	0.057	-0.791 ± 1.122	-0.991 ± 1.020	0.013 ^a
Lumbar spine						
Total area, cm ²	56.015 ± 5.102	56.496 ± 4.944	0.069	58.392 ± 4.640	59.438 ± 4.355	0.005 ^a
BMC, g	52.917 ± 9.297	54.097 ± 9.226	0.004 ^a	51.687 ± 8.611	53.175 ± 7.872	0.015 ^a
Reference aBMD, g/cm ²	0.988 ± 0.105	1.005 ± 0.103	—	0.957 ± 0.117	0.997 ± 0.119	—
aBMD, g/cm ²	0.951 ± 0.106	0.964 ± 0.107	0.001 ^a	0.870 ± 0.119	0.878 ± 0.107	0.386
Z score: BMD	-0.320 ± 1.088	-0.315 ± 1.094	0.938	-0.765 ± 1.083	-1.145 ± 0.936	0.002 ^a

Data are mean ± SD or median (interquartile range). Z scores: mean of Z scores of individual participants; age ranges: before L (11.9–18.4), after L (16.0–21.7), before CA (13.5–18.4), after CA (16.0–19.1), based on data from the NHANES (17) and Bone Mineral Density in Childhood Study (18) for the lumbar spine and hip, respectively.

Abbreviations: After L/CA, mean/median levels before start of CSH; Before L/CA, mean/median values before start of progestin.

^a $P < 0.05$.

^bThese rows represent reference values of age-matched peers of the same gender recorded at birth.

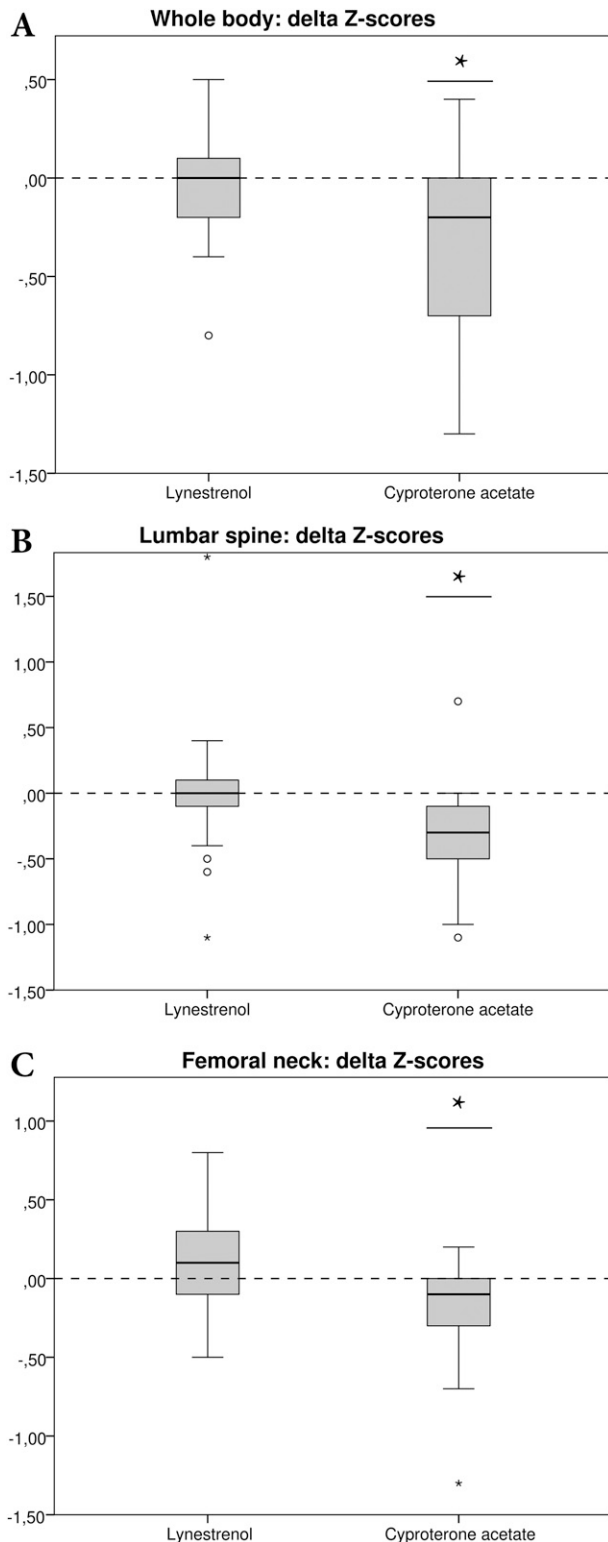


Figure 2. Changes in aBMD Z scores. Box and whisker plot of delta Z scores: aBMD of (A) whole body (17, 18), (B) lumbar spine (17), and (C) femoral neck (18). Boxes indicate the 25th and 75th percentiles; horizontal lines within the shaded boxes indicate the median. Whiskers indicate the minimum and maximum; mild and extreme outliers (circles and stars, respectively) are defined as 1.5 and 3 times the interquartile range below or above the 25th and 75th percentile, respectively. * $P < 0.05$.

Based on pQCT data, trabecular vBMD decreased at the distal radius and tibia, and Z scores of both total and trabecular vBMD at the radius decreased. Significant increases of the cortical BMC and vBMD were noted at both midshaft tibia and radius. SSIP Z scores at the midshaft radius decreased significantly. Cortical thickness and periosteal and endosteal circumferences of both radius and tibia did not change during treatment, resulting in significantly lower periosteal circumference Z scores in trans girls at the end of treatment. All pQCT data are represented in Supplemental Table 2.

Discussion

In our previous studies, we have shown that proandrogenic and antiandrogenic progestins, although not fully suppressing endogenous gonadotropin and sex steroid production, effectively reduce physical signs of puberty, such as menstruation in trans boys and erections and facial hair growth in trans girls (6, 7). In this study, we investigate whether these medications affect bone mass acquisition in comparison with age-matched peers and whether the induced hormonal changes are sufficiently potent to alter body composition and can therefore actually be considered a first step toward CSH treatment.

In trans boys taking L, lean mass and grip strength significantly increased. Similar effects on lean body mass and grip strength have been reported in adult trans men using androgens (20, 21). Compared with the results published by Ortega *et al.* (16) based on a large European cohort, the increase in grip strength in our study more than doubled the increase in healthy female peers and equaled the increase observed in male teens of similar age. These findings are corroborated by our pQCT results, showing an increase in the muscle area of the forearm, which was significantly greater than in age-matched female peers. Furthermore, when comparing the observed changes in waist and hip circumference with data obtained in a similar age group in a recent Dutch normative study of healthy pubertal boys and girls (22), our data suggest interference with physiological body development in trans boys, who had greater increases in waist and hip circumference. These data combined indicate an L-induced, androgenic effect on lean body mass and body musculature in trans boys and the development of a more masculine body shape.

Contrasting changes were observed in trans girls treated with CA (*i.e.*, lean body mass decreased, whereas fat mass increased), in line with findings in adult trans women using CA in combination with estrogens, who were found to have lower lean mass and higher fat mass than male control subjects (23). In our study, cross-sectional muscle area at the radius significantly decreased

in trans girls, accompanied by a stable grip strength. Compared with European peers, grip strength would be expected to increase during the same period of life in both sexes (16). As skeletal musculature in adolescents is normally still developing at this age, we consider our findings of relevance. Increased fat mass and decreased lean mass have also been observed in trans girls treated with GnRHa (3). We have previously reported, apart from a reduction in facial and body hair growth and sexual arousal, moderate breast development (B2/3) in one-third of trans girls treated with CA, which is not seen in adolescents using GnRHa (7). If confirmed in other studies, this may represent an additional benefit of CA over GnRHa in late-pubertal adolescents who already have fully developed (male) secondary sex characteristics.

Noteworthy is that all body composition changes in both groups occurred in <1 year of treatment and in the absence of CSH. We did not investigate whether adolescents themselves were aware of their bodies slightly changing toward the desired appearance, but the quantification of such self-perceived changes deserves to be the subject of future studies.

Trans boys showed similar increases as their age-matched female peers in aBMD and vBMD at all sites, and no major changes were seen in cortical thickness. A wealth of studies has evaluated the effects of progestin-only contraceptives, mostly depot medroxyprogesterone acetate, on bone. These studies reveal an overall negative but reversible impact on aBMD in adult women (24–27). In contrast, depot medroxyprogesterone acetate and progestin-only pills protect aBMD during lactation, which is considered to be a hypoestrogenic period [28, for review, see Curtis and Martins (29)].

In line with previous studies (4, 5), negative Z scores for aBMD were present in trans girls already at the start of treatment (29). Although not further investigated in this study, we hypothesize that reduced physical activity as compared with age-matched male peers may underlie this observation. Although vitamin D levels were suboptimal in trans girls, these levels did not differ from those of the trans boys Z in our study and from the general Belgian pediatric and adult population, who have been shown to have low vitamin D levels as well (30, 31).

In contrast to age-matched male peers, aBMD did not increase at any site in trans girls during the study period, and Z scores showed similar decreases at the lumbar spine and femoral neck, as has been reported for trans girls using GnRHa (4, 5). In addition, CA had a negative impact on vBMD of the trabecular bone of the distal radius. Z scores of the periosteal circumference decreased, indicating markedly less periosteal expansion as compared with their male peers. In contrast to the neutral effects of L on bone mass and geometry in trans boys, CA

treatment in trans girls appears to suppress both normal pubertal periosteal bone expansion as well as trabecular bone acquisition. We speculate that the reduced androgen levels seen in our adolescents taking CA underlie this observation. Indeed, it has been established that androgens are important regulators of periosteal bone apposition (32–36). SSIP at the midshaft radius increased in both patient groups; however, in trans girls, this increase was smaller than in their age-matched male peers. As SSIP strongly correlates with ultimate bone strength, this finding suggests a potentially higher future fracture risk in trans women (37, 38). Although the decline in markers of bone turnover was most pronounced in trans girls, a physiological decrease after Tanner B3/G3 has been reported in cross-sectional, population-based studies (39, 40). Therefore, these parameters were not helpful in documenting eventual treatment-induced changes.

Strengths of this prospective study are the relatively large cohort of investigated adolescents and the combination of DXA and pQCT technology, allowing a more detailed and mechanistic insight in trabecular and cortical bone changes as well as the evolution of body composition. A limitation of the study is the lack of a control group and, as a surrogate, comparison with several published cohorts of adolescents to calculate Z scores. It should be emphasized that in this study, a homogenous cohort of late-pubertal trans adolescents, with advanced endogenous pubertal hormone production and attendant physical and bone development at the start of the treatment, has been investigated. In this specific age group, the focus of treatment may have shifted from full suppression of pubertal hormones, which is achieved by GnRHa, toward alleviating the burden of pubertal signs and preparing the individual for CSH therapy. Studies of longer duration and randomized trials directly comparing GnRHa and progestin-induced changes in this specific population are required to define the clinical significance and impact of both treatment strategies on overall health and quality of life, both in the short and longer term.

In conclusion, treatment with proandrogenic and antiandrogenic progestins induces relevant changes in body composition and grip strength in line with gender identity in trans youth. These effects were most obvious in trans boys. In addition, similar to GnRHa treatment, CA interferes with pubertal bone development, despite general measures advised to promote bone health, such as a calcium-rich diet and routine vitamin D supplementation. This therefore warrants long-term follow-up studies on the evolution of bone strength-associated parameters, especially in trans girls, to assess whether this phenomenon is reversible after addition of CSH and

whether this would result in increased fracture rates in elderly individuals.

Acknowledgments

The authors thank Dr. Tom Fiers for assistance with the interpretation of laboratory techniques and reference values and all of the participants who allowed use of their data in the study.

Author Contributions: L.J.W.T. analyzed all data, including statistics, and drafted the manuscript. M. Craen, B.L., S.G., K.T., J.-M.K., S.V., G.T., and H.-G.Z. participated in data collection and interpretation of results. M. Cools participated in data collection, created the study design, and supervised the project and manuscript drafting. All authors critically revised the manuscript and approved the final version.

Correspondence and Reprint Requests: Martine Cools, MD, PhD, Princess Elisabeth Children's Hospital, Building 3K12D Room 319, C. Heymanslaan 10, 9000 Ghent, Belgium. E-mail: martine.cools@ugent.be.

Disclosure Summary: M.C. is a senior clinical investigator of the Research Foundation–Flanders. The remaining authors have nothing to disclose.

References

- Bonifacio HJ, Rosenthal SM. Gender variance and dysphoria in children and adolescents. *Pediatr Clin North Am*. 2015;62(4):1001–1016.
- Hembree WC, Cohen-Kettenis PT, Gooren L, Hannema SE, Meyer WJ, Murad MH, Rosenthal SM, Safer JD, Tangpricha V, T'Sjoen GG. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2017;102(11):3869–3903.
- Schagen SE, Cohen-Kettenis PT, Delemarre-van de Waal HA, Hannema SE. Efficacy and safety of gonadotropin-releasing hormone agonist treatment to suppress puberty in gender dysphoric adolescents. *J Sex Med*. 2016;13(7):1125–1132.
- Klink D, Caris M, Heijboer A, van Trotsenburg M, Rotteveel J. Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. *J Clin Endocrinol Metab*. 2015;100(2):E270–E275.
- Vlot MC, Klink DT, den Heijer M, Blankenstein MA, Rotteveel J, Heijboer AC. Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. *Bone*. 2017;95:11–19.
- Tack LJW, Craen M, Dhondt K, Vanden Bossche H, Laridaen J, Cools M. Consecutive lynestrenol and cross-sex hormone treatment in biological female adolescents with gender dysphoria: a retrospective analysis. *Biol Sex Differ*. 2016;7(1):14.
- Tack LJW, Heyse R, Craen M, Dhondt K, Bossche HV, Laridaen J, Cools M. Consecutive cyproterone acetate and estradiol treatment in late-pubertal transgender female adolescents. *J Sex Med*. 2017;14(5):747–757.
- Giltay EJ, Gooren LJG. Effects of sex steroid deprivation/administration on hair growth and skin sebum production in transsexual males and females. *J Clin Endocrinol Metab*. 2000;85(8):2913–2921.
- Gooren LJ, Giltay EJ, Bunck MC. Long-term treatment of transsexuals with cross-sex hormones: extensive personal experience. *J Clin Endocrinol Metab*. 2008;93(1):19–25.
- Wierckx K, Van Caenegem E, Schreiner T, Haraldsen I, Fisher AD, Toye K, Kaufman JM, T'Sjoen G. Cross-sex hormone therapy in trans persons is safe and effective at short-time follow-up: results from the European network for the investigation of gender incongruence [published correction appears in *J Sex Med*. 2016;13(4):732]. *J Sex Med*. 2014;11(8):1999–2011.
- Van Caenegem E, Wierckx K, Taes Y, Schreiner T, Vandewalle S, Toye K, Kaufman J-M, T'Sjoen G. Preservation of volumetric bone density and geometry in trans women during cross-sex hormonal therapy: a prospective observational study. *Osteoporos Int*. 2015;26(1):35–47.
- Wierckx K, Mueller S, Weyers S, Van Caenegem E, Roef G, Heylens G, T'Sjoen G. Long-term evaluation of cross-sex hormone treatment in transsexual persons. *J Sex Med*. 2012;9(10):2641–2651.
- Cools M, Craen M, T'Sjoen G. The endocrine treatment of transsexual adolescents. *Tijdschr Voor Geneeskunde*. 2010;12(1):28–33.
- Roelants M, Hauspie R, Hoppenbrouwers K. References for growth and pubertal development from birth to 21 years in Flanders, Belgium. *Ann Hum Biol*. 2009;36(6):680–694.
- Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. *J Clin Endocrinol Metab*. 1999;84(10):3666–3672.
- Ortega FB, Artero EG, Ruiz JR, España-Romero V, Jiménez-Pavón D, Vicente-Rodriguez G, Moreno LA, Manios Y, Béghin L, Ottevaere C, Ciarapica D, Sarri K, Dietrich S, Blair SN, Kersting M, Molnar D, González-Gross M, Gutiérrez A, Sjöström M, Castillo MJ; HELENA study. Physical fitness levels among European adolescents: the HELENA study. *Br J Sports Med*. 2011;45(1):20–29.
- Kelly TL, Wilson KE, Heymsfield SB. Dual energy X-Ray absorptiometry body composition reference values from NHANES. *PLoS One*. 2009;4(9):e7038.
- Kalkwarf HJ, Zemel BS, Gilsanz V, Lappe JM, Horlick M, Oberfield S, Mahboubi S, Fan B, Frederick MM, Winer K, Shepherd JA. The Bone Mineral Density in Childhood Study: bone mineral content and density according to age, sex, and race. *J Clin Endocrinol Metab*. 2007;92(6):2087–2099.
- Roggen I, Roelants M, Sioen I, Vandewalle S, De Henauf S, Goemaere S, Kaufman J-M, De Schepper J. Pediatric reference values for tibial trabecular bone mineral density and bone geometry parameters using peripheral quantitative computed tomography [published correction appears in *Calcif Tissue Int*. 2015;97(4):426–427]. *Calcif Tissue Int*. 2015;96(6):527–533.
- Van Caenegem E, Wierckx K, Taes Y, Schreiner T, Vandewalle S, Toye K, Lapauw B, Kaufman J-M, T'Sjoen G. Body composition, bone turnover, and bone mass in trans men during testosterone treatment: 1-year follow-up data from a prospective case-controlled study (ENIGI). *Eur J Endocrinol*. 2015;172(2):163–171.
- Mueller A, Haeberle L, Zollner H, Claassen T, Kronawitter D, Oppelt PG, Cupisti S, Beckmann MW, Ditttrich R. Effects of intramuscular testosterone undecanoate on body composition and bone mineral density in female-to-male transsexuals. *J Sex Med*. 2010;7(9):3190–3198.
- Fredriks AM, van Buuren S, Fekkes M, Verloove-Vanhorick SP, Wit JM. Are age references for waist circumference, hip circumference and waist-hip ratio in Dutch children useful in clinical practice? *Eur J Pediatr*. 2005;164(4):216–222.
- Lapauw B, Taes Y, Simoons S, Van Caenegem E, Weyers S, Goemaere S, Toye K, Kaufman JM, T'Sjoen GG. Body composition, volumetric and areal bone parameters in male-to-female transsexual persons. *Bone*. 2008;43(6):1016–1021.
- Paiva LC, Pinto-Neto AM, Faundes A. Bone density among long-term users of medroxyprogesterone acetate as a contraceptive. *Contraception*. 1998;58(6):351–355.
- Rosenberg L, Zhang Y, Constant D, Cooper D, Kalla AA, Micklesfield L, Hoffman M. Bone status after cessation of use of injectable progestin contraceptives. *Contraception*. 2007;76(6):425–431.

26. Lange HLH, Manos BE, Gothard MD, Rogers LK, Bonny AE. Bone Mineral Density and Weight Changes in Adolescents Randomized to 3 Doses of Depot Medroxyprogesterone Acetate. *J Pediatr Adolesc Gynecol*. 2017;30(2):169–175.
27. Modesto W, Bahamondes MV, Bahamondes L. Prevalence of low bone mass and osteoporosis in long-term users of the injectable contraceptive depot medroxyprogesterone acetate. *J Womens Health (Larchmt)*. 2015;24(8):636–640.
28. Costa ML, Cecatti JG, Krupa FG, Rehder PM, Sousa MH, Costa-Paiva L. Progestin-only contraception prevents bone loss in postpartum breastfeeding women. *Contraception*. 2012;85(4):374–380.
29. Curtis KM, Martins SL. Progestogen-only contraception and bone mineral density: a systematic review. *Contraception*. 2006;73(5):470–487.
30. Sioen I, Mouratidou T, Kaufman J-M, Bammann K, Michels N, Pigeot I, Vanaelst B, Vyncke K, De Henauw S; IDEFICS consortium. Determinants of vitamin D status in young children: results from the Belgian arm of the IDEFICS (Identification and Prevention of Dietary- and Lifestyle-Induced Health Effects in Children and Infants) Study. *Public Health Nutr*. 2012;15(6):1093–1099.
31. Moreno-Reyes R, Carpentier YA, Boelaert M, El Moumni K, Dufourny G, Bazelmans C, Levêque A, Gervy C, Goldman S. Vitamin D deficiency and hyperparathyroidism in relation to ethnicity: a cross-sectional survey in healthy adults. *Eur J Nutr*. 2009;48(1):31–37.
32. Vanderschueren D, Vandenput L, Boonen S, Lindberg MK, Bouillon R, Ohlsson C. Androgens and bone. *Endocr Rev*. 2004;25(3):389–425.
33. Bouillon R, Bex M, Vanderschueren D, Boonen S. Estrogens are essential for male pubertal periosteal bone expansion. *J Clin Endocrinol Metab*. 2004;89(12):6025–6029.
34. Taes Y, Lapauw B, Vandewalle S, Zmierzak H, Goemaere S, Vanderschueren D, Kaufman JM, T'Sjoen G. Estrogen-specific action on bone geometry and volumetric bone density: longitudinal observations in an adult with complete androgen insensitivity. *Bone*. 2009;45(2):392–397.
35. Lorentzon M, Swanson C, Andersson N, Mellström D, Ohlsson C. Free testosterone is a positive, whereas free estradiol is a negative, predictor of cortical bone size in young Swedish men: the GOOD study. *J Bone Miner Res*. 2005;20(8):1334–1341.
36. Lapauw BM, Taes Y, Bogaert V, Vanbillemont G, Goemaere S, Zmierzak H-G, De Bacquer D, Kaufman J-M. Serum estradiol is associated with volumetric BMD and modulates the impact of physical activity on bone size at the age of peak bone mass: a study in healthy male siblings. *J Bone Miner Res*. 2009;24(6):1075–1085.
37. Sheu Y, Zmuda JM, Boudreau RM, Petit MA, Ensrud KE, Bauer DC, Gordon CL, Orwoll ES, Cauley JA; Osteoporotic Fractures in Men MrOS Research Group. Bone strength measured by peripheral quantitative computed tomography and the risk of non-vertebral fractures: the osteoporotic fractures in men (MrOS) study. *J Bone Miner Res*. 2011;26(1):63–71.
38. Ferretti JL, Capozza RF, Zanchetta JR. Mechanical validation of a tomographic (pQCT) index for noninvasive estimation of rat femur bending strength. *Bone*. 1996;18(2):97–102.
39. Bayer M. Reference values of osteocalcin and procollagen type I N-propeptide plasma levels in a healthy Central European population aged 0-18 years. *Osteoporos Int*. 2014;25(2):729–736.
40. Rauchenzauner M, Schmid A, Heinz-Erian P, Kapelari K, Falkensammer G, Griesmacher A, Finkenstedt G, Högl W. Sex- and age-specific reference curves for serum markers of bone turnover in healthy children from 2 months to 18 years. *J Clin Endocrinol Metab*. 2007;92(2):443–449.



Testimony before the Senate Judiciary Committee

Riley Gaines

Advisor

Independent Women's Voice

Wednesday, June 21, 2023

My name is Riley Gaines, and I am an Advisor for Independent Women's Voice. Prior to joining IW, I was a student at the University of Kentucky where I was a member of UK's Women's Swim and Dive Team. I take pride that I finished my career as a 12x NCAA All-American, 5x SEC champion, SEC record holder in the 200 butterfly making me the 10th fastest American of all time in the event, and 2x Olympic trial qualifier.

On March 17th, 2022, my teammates and I, as well as female swimmers from universities around the country, were forced to compete against a biological male named Lia Thomas. Thomas was allowed to compete in the women's division after competing as a member of the University of Pennsylvania Men's Swim Team for three years as Will Thomas. We watched on the side of the pool as Thomas swam to a national title in the 500-yard freestyle, beating out the most impressive and accomplished female swimmers in the country, including Olympians and American record holders. (Previously, Thomas had been ranked in the 400ths — at best — in the men's division.)

The next day, I raced Thomas in the 200-yard freestyle. We tied. We ended up with the same time — down to the hundredth of a second.

DEFENDANT'S
EXHIBIT

34

Having only one trophy, the NCAA handed it to Thomas and told me I would go home empty-handed because Thomas needed to hold the trophy for photo purposes. I was shocked. I felt betrayed and belittled, reduced to a photo-op. But my feelings did not matter. What mattered to the NCAA were the feelings of a biological male.

In 1972, Congress enacted Title IX to end unjust sex discrimination in all aspects of education, including college athletics. But by allowing Thomas to displace female athletes — in the pool and on the podium — the NCAA intentionally and explicitly discriminated on the basis of sex. Although the NCAA claimed it acted in the name of “inclusion,” its policies, in fact, EXcluded female athletes, the very athletes whom Title IX was passed to protect.

But that is not all. In addition to being forced to give up our awards, our titles, and our opportunities, the NCAA forced me and my female swimmers to share a locker room with Thomas, a 6’4” 22-year-old male equipped with (and exposing) male genitalia. Let me be clear: we were not forewarned. We were not asked for our consent. And we did not give our consent. In swimming locker rooms you’re undressing quickly in front of one another, fully naked. I was forced to take off my swimsuit in front of a man. If nothing else, I truly hope you can see how this is a violation of our right to privacy and how some of us have felt uncomfortable, embarrassed, and even traumatized by the experience.

I know that I do not speak for everyone who competed against Lia Thomas. But I do speak for many. I can attest to the tears that I witnessed from finishers who missed being named an All-American by one place. I can attest to the extreme discomfort in the locker room from 18-year-old girls exposed

to male body parts and having to undress with a male watching in the same room. I can attest to the anger and frustration expressed by girls who had worked so hard and sacrificed so much to get to this moment, only to have to compete in what felt like a farce. And I can attest to the fact that, around the country, female athletes who protested the inclusion of Lia Thomas in the women's division were threatened, intimidated, and emotionally blackmailed into silence and submission.

Unfortunately, our experiences are not unique. The number of female athletes who have been denied opportunities, traumatized, or hurt by policies that claim to promote "inclusion" is growing at an alarming rate. I have heard from female athletes who have been seriously injured – one with permanent injuries that will plague her for the rest of her life – because she was forced to compete against a much physically stronger man. This is simply unacceptable and the integrity of women's sports is lost. It is unfair, and it is discriminatory. And it must stop.

Women's rights to privacy and single-sex spaces and opportunities are being encroached upon — sports, sororities, locker and dorm rooms, shelters, and prisons.

Some have tried to tar those of us speaking up for women's safety, security, and opportunities as "transphobic" or "bigoted." This is untrue.

I've heard from gay, lesbian, and trans-identifying Americans that agree females should not be asked to step aside to make room for male-bodied individuals — no matter how they self identify.

Defending women's rights is not anti-anyone. Believing in biology is not bigoted. And following the science that there are two sexes and that there are very real and important differences between the two sexes is not hateful, it's fact.

I'll end with a quote from tennis legend Martina Navratilova, "[T]here will always be significant numbers of boys and men who would beat the best girls and women in head-to-head competition. Claims to the contrary are simply a denial of science."

I appreciate the opportunity to share my experience and hope you really, truly, heard my story and consider the facts, the science, and the growing stories. Thank you.