EXHIBIT 57

Endocrine Reviews, 2021, Vol. 42, No. 3, 219–258
doi:10.1210/endrev/bnaa034
Scientific Statement

Scientific Statement

Considering Sex as a Biological Variable in Basic and Clinical Studies: An Endocrine Society Scientific Statement

Aditi Bhargava,[1,2] Arthur P. Arnold,[3] Debra A. Bangasser,[4] Kate M. Denton,[5] Arpana Gupta,[6] Lucinda M. Hilliard Krause,[5] Emeran A. Mayer,[6] Margaret McCarthy,[7] Walter L. Miller,[1,8] Armin Raznahan,[9] and Ragini Verma[10]

[1]Center for Reproductive Sciences and [2]Department of Obstetrics and Gynecology, University of California, San Francisco, CA 94143, USA; [3]Department of Integrative Biology & Physiology, University of California, Los Angeles, Los Angeles, CA 90095, USA; [4]Department of Psychology and Neuroscience Program, Temple University, Philadelphia, PA 19122, USA; [5]Cardiovascular Disease Program, Monash Biomedicine Discovery Institute and Department of Physiology, Monash University, Clayton, Victoria, 3800, Australia; [6]G. Oppenheimer Center for Neurobiology of Stress and Resilience, Division of Digestive Diseases, University of California, Los Angeles, Los Angeles, CA 90095-7378, USA; [7]Department of Pharmacology and Program in Neuroscience, University of Maryland School of Medicine, Baltimore, MD 21201, USA; [8]Department of Pediatrics, University of California, San Francisco, CA 94143, USA; [9]Section on Developmental Neurogenomics, Human Genetics Branch, National Institutes of Mental Health, Intramural Research Program, Bethesda, MD 20892, USA; and [10]Diffusion and Connectomics In Precision Healthcare Research (DiCIPHR) lab, Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

ORCiD number: 0000-0003-1334-0517 (A. Bhargava).

Abbreviations: ACTH, adrenocorticotropic hormone; AT_2R, angiotensin type 2 receptor; BMI, body mass index; cAMP, cyclic adenosine monophosphate; CKD, chronic kidney disease; CRF, corticotropin-releasing factor; CVD, cardiovascular disease; dMRI, diffusion magnetic resonance imaging; fMRI, functional magnetic resonance imaging; FCG, Four Core Genotypes (model); GMV, gray matter volume; GPCR, G-protein coupled receptor; HPA, hypothalamic-pituitary-adrenal; KYN, kynurenine; LC, locus coeruleus; MIH, Müllerian inhibitory hormone; PAR, pseudoautosomal region; PKA, protein kinase A; PTSD, posttraumatic stress disorder; RAAS, renin-angiotensin-aldosterone system; rs-fMRI, resting state functional magnetic resonance imaging; sMRI, structural magnetic resonance imaging; UCN, urocortin.

Received: 22 December 2020; First Published Online: 11 March 2021; Corrected and Typeset: 11 March 2021.

Abstract

In May 2014, the National Institutes of Health (NIH) stated its intent to "require applicants to consider sex as a biological variable (SABV) in the design and analysis of NIH-funded research involving animals and cells." Since then, proposed research plans that include animals routinely state that both sexes/genders will be used; however, in many instances, researchers and reviewers are at a loss about the issue of sex differences. Moreover, the terms *sex* and *gender* are used interchangeably by many researchers,

ISSN Print: 0163-769X
ISSN Online: 1945-7189
Printed: in USA

further complicating the issue. In addition, the sex or gender of the researcher might influence study outcomes, especially those concerning behavioral studies, in both animals and humans. The act of observation may change the outcome (the "observer effect") and any experimental manipulation, no matter how well-controlled, is subject to it. This is nowhere more applicable than in physiology and behavior. The sex of established cultured cell lines is another issue, in addition to aneuploidy; chromosomal numbers can change as cells are passaged. Additionally, culture medium contains steroids, growth hormone, and insulin that might influence expression of various genes. These issues often are not taken into account, determined, or even considered. Issues pertaining to the "sex" of cultured cells are beyond the scope of this Statement. However, we will discuss the factors that influence sex and gender in both basic research (that using animal models) and clinical research (that involving human subjects), as well as in some areas of science where sex differences are routinely studied. Sex differences in baseline physiology and associated mechanisms form the foundation for understanding sex differences in diseases pathology, treatments, and outcomes. The purpose of this Statement is to highlight lessons learned, caveats, and what to consider when evaluating data pertaining to sex differences, using 3 areas of research as examples; it is not intended to serve as a guideline for research design.

Key Words: brain-gut, cardiovascular disease, chromosome complement, gender, sex differences, steroid hormones

Sex is an important biological variable that must be considered in the design and analysis of human and animal research. The terms *sex* and *gender* should not be used interchangeably. Sex is dichotomous, with sex determination in the fertilized zygote stemming from unequal expression of sex chromosomal genes. By contrast, gender includes perception of the individual as male, female, or other, both by the individual and by society; both humans and animals have sex, but only humans have gender. Both sexes produce estrogens, androgens, and progestins; there are no male- or female-specific sex hormones, *per se*, although these steroids are present in substantially different levels in males and females. Sex differences are caused by 3 major factors—sex hormones, genes, and environment. To understand disease mechanisms and exploit sex differences in protection or exacerbation of diseases, one needs to determine the relative contribution of factors, including observer effect (1), causing sex differences. Here—using 3 broad research areas as examples—the roles of sex differences in brain anatomy, brain-gut axis, and cardiovascular disease are discussed. Contemporary brain imaging methods show age- and sex-related differences in brain size, global and regional gray matter volume, white matter connectivity, and neuroanatomic regulation of appetite and satiety; while these differences are seen in large population-based studies, there is tremendous individual overlap, but such group-level findings do not inform findings, physiology, or pathology at the individual level. Sex differences in disorders of the brain-gut axis, obesity, type 2 diabetes, and metabolic syndrome are caused by differential actions of brain-gut peptide and steroid hormones. The activation, signaling, and pharmacotherapy responses of the components of the hypothalamic-pituitary-adrenal (HPA) axis differ between the sexes. Heart and kidney functions are linked. Age, hormones, and sex biases seen in cardiovascular and chronic kidney diseases also differentially influence pharmacologic responses in men and women. Thus, sex differences pervade biology and medicine, and while not discussed in this Statement, must be considered in virtually all areas of biomedical research.

Section I

Sex Versus Gender

Much of the American public is surprisingly prudish about the word *sex*; it has now become commonplace to use the seemingly more genteel term *gender* when one really means *sex*. In *Moritz v Commissioner of Internal Revenue* (469 F. 2d 466 [1972]), Ruth Bader Ginsburg (subsequently, The Honorable Ruth Bader Ginsburg) argued against discrimination "on the basis of sex" not "on the basis of gender," thus clearly, knowledgeably, and presciently understanding that "sex" does not equal "gender." In a decision 48 years later (*Bostock v Clayton County*, 590 US, decided June 15, 2020), the United States Supreme Court separately ruled against discrimination on the basis of gender. *Gender* is often misused as a synonym for *sex*—for example, when filling out forms for various activities, we are routinely

asked to check a box labeled "gender," but the only available options are boxes labeled "M" and "F." But *sex* is not the same thing as *gender* and using these terms as equivalents obfuscates differences that are real and important in society in general and biomedical research in particular.

Biological Sex: The Definition of Male and Female

Sex is a biological concept. Asexual reproduction (cloning) is routine in microorganisms and some plants, but most vertebrates and all mammals have 2 distinct sexes. Even single-cell organisms have "mating types" to facilitate sexual reproduction. Only cells belonging to different mating types can fuse together to reproduce sexually (2, 3). Sexual reproduction allows for exchange of genetic information and promotes genetic diversity. The classical biological definition of the **2 sexes** is that females have ovaries and make larger female gametes (eggs), whereas males have testes and make smaller male gametes (sperm); the 2 gametes fertilize to form the zygote, which has the potential to become a new individual. The advantage of this simple definition is first that it can be applied universally to any species of sexually reproducing organism. Second, it is a bedrock concept of evolution, because selection of traits may differ in the 2 sexes. Thirdly, the definition can be extended to the ovaries and testes, and in this way the categories—female and male—can be applied also to individuals who have gonads but do not make gametes.

In mammals, numerous sexual traits (gonads, genitalia, etc) that typically differ in males and females are tightly linked to each other because one characteristic leads to sex differences in other traits. The type of gonads is controlled by the presence of XX or XY chromosomes, and gonadal secretions in turn regulate formation of female or male reproductive tissues, and characteristics that differ in typical males or females. These characteristics include external genitalia, uterus and oviducts, sperm ducts, and secondary sexual characteristics such as facial hair and pitch of voice. However, many people cannot make either eggs or sperm, yet are recognized as female or male based on other physical characteristics; people who do not have either ovaries or testes are rare. For individuals that possess a combination of male- and female-typical characteristics, these clusters of traits are sufficient to classify most individuals as either biologically male or female. For example, a person with testes and a penis, who cannot make sperm, is usually classified as a biological male, as long as the person does not possess female features such as a vagina, ovaries, or uterus. Based on evidence presented, to define male and female individuals in general society, we expand the defining characteristics of sex to include nongonadal traits, as well as classical gonadal traits.

A simple biological definition of male and female, satisfactory to all people, is elusive. In human societies, the terms *female* and *male* can have several meanings, as they refer both to a person's biological sex and to their social roles. Most people learn to discriminate males and females from an early age, but often not based on biological traits (4). For example, behaviors such as pair-bonding, sexual activity, offspring defense and care, and mate/partner selection (5) involve complex interplay between sex steroid hormones and peptide hormones (oxytocin and arginine vasopressin); these behaviors are encouraged differently in women and men, which influences their role in the society and culture in which they live to behave as "females" or "males." While these factors have little impact on their biological sex, they can have profoundly different outcomes in the behavior and health of an individual. Biological sex is dichotomous because of the different roles of each sex in reproduction. For scientific research, it is important to define biological sex and distinguish it from other meanings.

Sex Chromosomes and Biological Sex Determination

Among mammals and many other taxa, males are characterized as the heterogametic sex (6), having 2 different sex chromosomes, X and Y, whereas females are homogametic (XX). By contrast birds, many reptiles, and some other organisms have Z and W chromosomes (7). In these organisms, the female is the heterogametic sex (ZW) and males are homogametic (ZZ). Some adult fish and reptiles can also change sex in response to environmental factors (8, 9), and even the adult mouse gonad can undergo partial sex reversal when specific genes are deleted (10, 11). Human biological sex is often assessed by examining the individual's complement of sex chromosomes as determined by karyotypic analysis: males are XY and females are XX. Karyotypic sex is actually a surrogate for genetic sex, determined by the presence of the *SRY* gene on the Y chromosome (12, 13). However, karyotypic analysis may be misleading, as there are well-described 46,XX males (with testes). Most of these individuals carry a short segment of the Y chromosome that includes *SRY* transferred to an X chromosome, but up to 10% lack an *SRY* gene (14, 15). Similarly, there are 46,XY females, who have *SRY* but also have a duplication of *DAX1* (*d*osage-sensitive sex reversal, *a*drenal hypoplasia critical region, on chromosome *X*, gene 1) (16).

Sex Determination and Sex Differentiation

In mammals, sex determination begins with the inheritance of XX or XY chromosomes, which are the only factors that are different in XX and XY zygotes. Thus, all phenotypic sex differences, including gonadal development, stem originally from the unequal effects of XX and XY

sex chromosomes. Phenotypic sex differences develop in XX and XY embryos as soon as transcription begins. The categories of X and Y genes that are unequally represented or expressed in male and female mammalian zygotes, which could cause phenotypic sex differences, fall into 3 main categories (17).

1. *Y genes causing male-specific effects.* These Y-linked genes do not have homologous genes on the X chromosome. The most important Y-linked gene is *SRY*, the testis-determining gene, which encodes the SRY transcription factor expressed during embryonic life in the bipotential gonadal ridge; SRY activates downstream autosomal genes such as *SOX9* to cause formation of a testis (18). In the absence of *SRY*, autosomal and X chromosome genes (*WNT-4*, *DAX-1*, *FOXL2*, *COUP-TFII*, and *RSPO1*) are activated to cause formation of an ovary (19-22). Both testicular and ovarian development are subject to active genetic regulation (12, 13, 16). Pathways downstream of *SRY* inhibit ovary-determining pathways, and ovary-determining pathways also inhibit pathways for testis development. Once the testes form, they secrete sex hormones that act widely throughout the body to cause male differentiation of nongonadal tissues. Other Y genes also have male-specific effects (for example, those required for spermatogenesis) (23, 24).
2. *X gene dosage or parental imprint.* Because XX nongermline cells inactivate one X chromosome (25, 26), it was long thought that both XX and XY cells have only one active X chromosome, with little inherent difference in expression related to the number of X chromosomes. The inactivated regions of the X chromosome are "coated" with large noncoding RNA transcribed from the X-inactive specific transcript (*XIST*) gene, part of the XIC (X inactivation center) located on Xq13 (27, 28). But some genes escape X inactivation (termed as *X escapees*), and therefore are expressed more in XX than XY cells, resulting in imbalance or incomplete dosage compensation (29). About 23% of human X-linked genes are more abundantly expressed in XX cells than XY cells in many tissues (30, 31). Recent evidence from mouse studies suggests that the inherent male-female difference in expression of X genes leads to significant sex differences in disease phenotypes. For example, sex differences in placental *Ogt* expression are associated with sex differences in prenatal vulnerability to stress (32). X escapee *Kdm6a*, a histone demethylase, contributes to sex differences in mouse models of bladder cancer (33), autoimmune disease (34), and Alzheimer disease (35). Similarly, variations in human *KDM6A* are associated with prognosis of bladder cancer or cognitive decline in female patients (33). The dose of another X escapee histone demethylase, *Kdm5c*, contributes to sex differences in adiposity and body weight in mice, and variations in *KDM5C* in humans are associated with body mass (36).

   Sex differences may also arise from genes in the pseudoautosomal regions (PARs) of the sex chromosomes, small regions of sequence similarity on the X and Y chromosomes that allow for X and Y chromosome pairing during meiosis. Both XX and XY cells have 2 PARs, implying equivalent effects of XX and XY PARs. Paradoxically, the process of X inactivation appears to spill over into the PAR and reduce expression on one X chromosome only in XX cells, leading to greater expression of PAR genes in XY cells compared to XX cells in the human transcriptome (30). A third potential source of X-linked imbalance stems from parentally imprinted genes in XX cells, which have one X chromosome from each parent and thus are influenced by any imprint on X genes from either parent. XY cells only receive imprints from the mother, and thus differ phenotypically from XX cells (37).
3. *XX mosaicism.* Female mammals are a mosaic of cells of 2 types: those expressing the X chromosome from the father (Xp), or from the mother (Xm) because of X inactivation (25). In contrast, XY individuals will lack this diversity within cell types in each organ because only one X (Xm) chromosome and only the maternal imprint of X genes will be expressed in each cell. The mosaicism in females means that in genetically diverse populations, the effects of disease-promoting X-linked alleles, inherited from one parent, will be muted in XX cells because half of the cells will have a different allele (38), and genomic imprints from each parent will only be expressed in half of the cells. In general, XX tissues are thought to have less extreme phenotypes than XY tissues, because the effects of extremely deleterious or beneficial alleles or imprints are buffered by the diversity of X alleles and imprints. For example, hemophilia A and hemophilia B (clotting factor VIII and IX deficiencies, respectively), are X-linked diseases that affect men, whereas most women are asymptomatic carriers.

Sexual Differentiation Caused by Gonadal and Nongonadal Hormones

In mammals, the process of reproductive system development requires the action of hormones (peptide/gonadotropins and steroids) from the pituitary gland, the adrenal cortex, and the gonads. Testicular development leads to secretion of Müllerian inhibitory hormone (MIH, also termed anti-Müllerian hormone, AMH), a glycopeptide, and testosterone, which affects many sex differences in nongonadal tissues (39). In contrast to the fetal testis, the fetal ovary makes minimal steroid hormones

(40), and ovarian function is not needed for development of the female reproductive system, as evidenced by the normal female anatomy of individuals with Turner syndrome, who have 45,X gonadal dysgenesis. The pioneering work of Alfred Jost suggested that 2 classes of testicular hormones are involved in sexual differentiation. First, testicular androgens drive the differentiation of the fetal external genitalia from female morphology to that of the male and are required for the differentiation of embryonic Wolffian ducts into male internal reproductive structures (41, 42). Androgens, secreted by Leydig cells, are required for the differentiation of embryonic Wolffian ducts into male internal reproductive structures (epididymis, vas deferens, ejaculatory ducts, prostate, and seminal vesicles), and drive the differentiation of the undifferentiated external genitalia toward male morphology. Second, the testis produces locally acting MIH that causes involution of the Müllerian ducts, which would otherwise develop into the fallopian tubes, uterus, and cervix (43, 44).

It was long thought that only the involution of the Müllerian ducts was an active process, with the Wolffian ducts simply involuting in the absence of androgens. Recent evidence from mice indicates that Wolffian involution is also an active process controlled by the transcription factor COUP-TFII (22, 45), but the nature of any factors stimulating COUP-TFII remains unknown (22). Some aspects of gonadal differentiation are active throughout life, preventing ovarian follicle cells from transdifferentiating into "testis-like" cells (11). MIH is secreted by Sertoli cells and androgenic steroid hormones, usually testosterone, are secreted by Leydig cells. Testosterone and its more potent derivative dihydrotestosterone are responsible for the development of the male external genitalia (46). Androgens from adrenal glands and alternative pathway androgen biosynthesis in the human placenta can influence virilization of the developing fetus (47, 48). The adrenals of adult primates also produce abundant androgens, profoundly influencing phenotypes, so that not all sex steroids are gonadal (see Boxes 1 and 2). Although the term *sexual differentiation* is usually applied to the development of sex differences in genitalia and other organs such as the brain in the growing fetus; sex differences also occur later in life during the mini-puberty of infancy (49), puberty, the female menstrual cycle, menopause in women, and andropause in men. The actions of gonadal and nongonadal hormones as well as sex and autosomal chromosome gene products in adult people causes many sex differences in health and disease.

Influence of Gonadal Steroid Hormones and Nongonadal Hormones in Brain Development

Differentiation of the brain by gonadal hormones is implemented during a restricted critical window, which is operationally defined by the onset of copious androgen

Box 1. Steroidogenesis in gonadal and nongonadal tissues

All biologically active sex steroids, whether gonadal or nongonadal in origin, are derived from cholesterol by the process of steroidogenesis. Two steroidogenic steps must be considered (for details see (50)). First, the cholesterol side-chain cleavage enzyme, P450scc (CYP11A1) initiates steroidogenesis by converting cholesterol to pregnenolone; expression of P450scc renders a tissue "steroidogenic," that is, able to make steroids de novo (51). The gonads, adrenals, and placenta express abundant P450scc and produce the familiar circulating endocrine steroids, but the brain, skin, and some other organs also express low levels of P450scc and produce steroids involved in paracrine actions. Brain steroidogenesis has been studied mainly in fetal rodents, with little information in other systems (52). Many nonsteroidogenic tissues (liver, kidney, fat, breast, heart) do not express P450scc but express other steroidogenic enzymes that modify steroids taken up from the circulation. Fat and breast express CYP19A1 (aromatase), permitting local production of estradiol from circulating 19-carbon (C19) steroids; this estradiol is important in breast cancer but is not a gonadal steroid. Similarly, prostate and genital skin express several enzymes leading to dihydrotestosterone, accounting for the failure of "androgen deprivation therapy" by gonadectomy in prostate cancer. Not all gonadal steroids are sex steroids, as both the ovary and testis secrete some "upstream" steroids that are precursors of the classic sex steroids. For example, dehydroepiandrosterone (DHEA) does not bind to sex steroid receptors, but it can be converted into testosterone and estrone. Second, synthesis of all sex steroids requires P450c17 (CYP17A1), which catalyzes 17α-hydroxylation and the 17,20 lyase activity that changes 21-carbon steroids to C19 precursors of androgens and estrogens. P450c17 is abundantly expressed in the gonads of all vertebrates and in the adrenals of most vertebrates other than rodents, but the rodent *Cyp17A1* gene is silenced by tissue-specific methylation (53). Consequently, rodents make only miniscule amounts of adrenal C19 steroids and also use corticosterone instead of cortisol as their glucocorticoid. In most mammals, P450c17 has low 17,20 lyase activity, so that their adrenals produce rather small amounts of C19 steroids, but primate P450c17 has abundant 17,20 lyase activity, generating abundant C19 androgen precursors (DHEA, DHEA-sulfate, androstenedione) (47, 48). Furthermore, production of these C19 steroids proceeds by different pathways in rodents and primates: primates favor the "Δ5 pathway," through DHEA, whereas rodents favor the "Δ4 pathway" through 17OH-progesterone (17OHP) (50). Primate adrenals also produce a true androgen, 11-keto-testosterone (54), profoundly influencing phenotypes (apocrine odor; female sexual hair). Thus, not all sex steroids are gonadal: ~ 50% of the circulating androgens in adult women are of adrenal origin.

Box 2. Gonadectomy and sex steroids

Many animal studies employ gonadectomy to eliminate the actions of sex steroids (estrogens, androgens, progestins). If using this approach, the investigator must consider whether nongonadal tissues will produce sufficient sex steroids to influence the study. The gonads produce most but not all circulating sex steroids; furthermore, some tissues produce steroids that act locally and do not enter the circulation, hence absence of a measurable steroids in blood does not ensure absence of its action in the target tissue. Both sexes produce all steroids and their metabolites, hence there are no male- or female-specific sex hormones, *per se*. In male mammals, testosterone release is highly pulsatile in nature (49, 55) and in laboratory mice, strain-dependent variations in androgen levels are reported (56). In female rodents, circulating levels of estradiol, testosterone, and DHT are highest in proestrus phase; a comprehensive analyses of sex steroids in intact and gonadectomized rodents can be found elsewhere (57). Circulating concentrations of testosterone in adult women are similar to those of boys in early puberty, and estradiol concentrations in men are similar to those in mid-cycle women, but the tenfold higher concentrations of testosterone obscure its effects. Rodents are widely used in research, but they differ from primates in several important aspects of steroidogenesis (see Box 1), and hence must be used with caution in studies seeking to model aspects of human physiology that might be influenced by steroids. These differences include: (i) In humans, substantial amounts of circulating sex steroids are bound to sex hormone–binding globulin (SHBG), whereas this carrier protein is not present in rodent circulation (58). (ii) Dehydroepiandrosterone (DHEA) and androstenedione, 19-carbon (C19) precursors for testosterone and estrone, that do not bind to sex steroid receptors, are secreted from the adrenal glands, the ovary and testis in humans, but not rodents (59). Thus, not all gonadal steroids are sex steroids. (iii) The rodent ovarian corpus luteum produces progesterone throughout pregnancy but in human pregnancy the corpus luteum involutes early in the second trimester, after which the placenta produces the progesterone needed to suppress uterine contractility, permitting term pregnancy. (iv) Adrenal-specific methylation of rodent *Cyp17A1* prohibits their adrenal synthesis of C19 precursors of sex steroids; however, changes in methylation status can occur under conditions of pathology. (v) As a further consequence of adrenal *Cyp17A1* methylation, rodents utilize corticosterone as their glucocorticoid, whereas almost all other vertebrates use cortisol. (vi) Rodent adrenals use high-density lipoproteins (HDL) taken up via scavenger receptor B1 (SRB1), as their principal source of cholesterol for steroidogenesis, whereas primates use low-density lipoproteins (LDL) taken up by receptor-mediated endocytosis. (vii) Several genes encoding steroidogenic enzymes are duplicated; rodents and primates differ in which copy(ies) of these genes are expressed: *CYP21; HSD3B, HSD17B, AKR1-3*. Such differences may affect laboratory results in unanticipated fashions. (viii) In rodents, nonsteroidogenic tissues such as the gut, liver, kidney, fat, breast, heart, thymus, skin, and the placenta have all been shown to make steroids. Thus, gonadectomy may eliminate most, but not all, circulating sex steroids, depending on the species being studied and may not reveal much about the paracrine effects of sex steroids present in the tissue(s) under investigation. Nonetheless, gonadectomy is an invaluable research tool that helps unequivocally confirm the influence of gonadal hormones in sex differences.

production from the fetal testis. Human fetal androgen production begins at 8 to 10 weeks postconception and in rodents is closer to parturition, at embryonic days 16 to 18, with birth following 2 to 4 days later. An important effect of this androgen surge is to masculinize the rodent brain. Steady but pulsatile release of the gonadotropins luteinizing hormone and follicle stimulating hormone from the pituitary gland support continuous steroidogenesis and production of sperm (60). In female rodents, the feminization of the brain proceeds in the absence of exposure to high levels of androgens or their aromatized byproducts, estrogens, a developmental strategy highly analogous to that used for masculinization of the gonads, reproductive tract, and secondary sexual characteristics, with the exception that estrogens are actively downregulated in male rodents. In human females, gonadotropins from the pituitary gland regulate ova development, induction of ovulation, and stimulation of estradiol and progesterone from the ovaries (49). An important feature of this developmental strategy is the existence of a sensitive period in female rodents (61). Male rodents must be exposed to high levels of androgens during the critical period; if exposure occurs too early or too late it will be ineffective at inducing masculinization. However, females are also sensitive to androgens during a restricted period of development, hence a sensitive period in rodents. In males, the critical period closes shortly after androgen exposure because the cellular and molecular processes of masculinization have been initiated and cannot be reversed; the train has left the station. In both primates and rodents this process is largely prenatal, but female rodents remain sensitive to androgen exposure into the first postnatal week. Injecting a newborn female rodent with androgens will initiate the process of masculinization, thus she is still sensitive. After the first week, the feminization process cannot be overridden by androgens and thus the sensitive period has closed. The existence of the sensitive period in females is useful as a research tool—it is important in understanding the potential impact of exposure to endocrine-disrupting compounds or other cellular agents of masculinization that act in an analogous manner to androgen exposure in modulating female brain development. There is evidence for a later sensitive

period for brain feminization mediated by small increases in estrogens (62); this topic warrants further investigation. The closing of the sensitive period in primates, especially humans, remains poorly understood, but it appears to end prenatally, similar to the critical period in rodents. The sources of androgens that females can be exposed to during the sensitive period include from: (i) experimental interventions; (ii) male littermates in animals; (iii) or human adrenals carrying genetic mutations in the steroidogenic pathway (as in congenital adrenal hyperplasia).

Given that the critical and sensitive periods for sexual differentiation are defined by the production and response to gonadal steroids, it is not surprising that steroids are the primary drivers of developmental origins of sex differences in brain (and probably other tissues) and behavior. But how do steroids achieve this? The first step in any investigation is often is to identify the active steroid metabolite(s). In rodents, circulating fetal testicular testosterone enters the fetal brain where it can serve as a direct precursor for estradiol synthesis via aromatase (*Cyp19A1*) (see Box 1). Fetal and adult neurons can aromatize testosterone to estradiol in a nonrandom distribution: neurons of the hypothalamus, preoptic area, and amygdala are particularly active for local estradiol synthesis, whereas the hippocampus and parts of the cortex, midbrain, and spinal cord are also active at a lower level (63). For most reproductive endpoints, it is the local actions of estradiol that drive neural phenotype toward masculinization, which to some seems counterintuitive, given that estradiol is so often referred to as a "female" hormone (64), and further highlights that it is impossible to completely eliminate the effects of sex steroids, especially in the brain, by simple gonadectomy (see Box 2). Developing rodent embryos sequester maternal estrogens by binding to circulating alpha-fetoprotein, which is present only during the critical/sensitive period; when it is genetically deleted, all the offspring are masculinized (65). However, in humans, sex hormone–binding protein, not alpha-fetoprotein, is the major serum glycoprotein that binds androgens and estrogens with an undetermined role in fetal sexual development (66, 67).

In rodents, there is abundant evidence that gonadal androgens are metabolized to estrogens in the brain and mediate "masculinizing" effects on the brain; similar evidence in primates is limited. In primates, the principal masculinizing agents are androgens, not estrogens, and although there is alpha-fetoprotein present in fetal circulation, it has a weak binding affinity for estradiol (68), and instead it plays a much broader role in brain and body development (69). The conclusion of no strong role for estrogens in humans is based on individuals with dysfunctional aromatase or androgen receptors. Males lacking aromatase still identify as men, while XY individuals with complete androgen insensitivity identify as women (70). The disparity between the principal differentiating hormones in primates versus rodents suggests that findings may not be easily extrapolated, and it is important to specify both the hormone and species under investigation. To discern whether the biological basis of sexual differentiation of brain and behavior differs between primates and rodents, one needs to identify mechanisms by which steroids transduce signals to modify the trajectory of the nervous system. While those mechanisms are incompletely understood, a few general principles are clear. First, there is no unified mechanism that applies broadly across the brain, with the exception that androgens and estrogens are the primary drivers of masculinization during a restricted developmental window. Similar masculinizing effects of testicular androgens may also occur during puberty (71). Second, all aspects of neural development are capable of being "organized" or programmed by sex steroids. This includes cell genesis, migration, myelination, dendritic and axonal growth and branching, synapse formation, synapse elimination, and neurochemical differentiation. Effects are not limited to neurons, with both astrocytes and microglia also exhibiting morphological sex differences. Third, each discrete brain region, nucleus, or subnucleus appears to have unique mechanisms of cellular masculinization. In some brain regions, such as the preoptic area, there are multiple separate mechanisms at play simultaneously. Sex steroids act in both paracrine and endocrine manners to influence structural development and function (72, 73).

Biological Basis of Diversity in Sexual/Gender Development and Orientation

Given the complexities of the biology of sexual determination and differentiation, it is not surprising that there are dozens of examples of variations or errors in these pathways associated with genetic mutations that are now well known to endocrinologists and geneticists (74); in medicine, these situations are generally termed *disorders of sexual development* (DSD) or *differences in sexual development* (75). DSD includes genetic disorders in the sexual determination pathway (76), disorders of steroidogenesis (50, 77), disorders of steroid hormone action, especially androgen insensitivity syndrome (78), and less well-defined "developmental field defects" (79), such as Mayer–Rokitansky–Küster–Hauser syndrome (80). The study of genes and factors underlying DSD and the diagnosis and management of the various forms of DSD is a complex and rapidly evolving area of endocrinology: clinical management is complex (81) and requires both contemporary molecular genetics (82) and well-integrated interdisciplinary care (83).

Gender includes perception of the individual as male, female, or other, both by the individual and by society. *Gender identity* is a psychological concept that refers to an individual's self-perception; while associations between gender identity, neuroanatomic, genetic, and hormone levels exist, a clear causative biological underpinning of gender identity remains to be demonstrated. Both animals and human beings have biological sex, but only humans have evident self-awareness that allows them to express gender; self-awareness in animals has not been investigated in this context. Gender also includes differences that males and females experience in their social and physical environments, which can have differentiating effects on the sexes. Human social environments are poorly modeled in laboratory animals and thus animal studies are usually limited to addressing sex differences. For centuries, the concept of male and female did not distinguish between biological sex differences and those caused by consistent differences in the environments. Thus *sex differences* are those caused by biological factors, whereas *gender differences* reflect a complex interplay of psychological, environmental, cultural, and biological factors (Fig. 1).

At birth, individuals are assigned a sex or gender ("natal gender"), almost always based on the appearance of the external genitalia. In most individuals, the various biological determinants of sex are consistent with one another, and this biological sex is also consistent with the individual's self-perception—the sex and gender are concordant. However, a substantial minority of people who do not have DSD have some degree of variation in their self-perception of their gender, which may differ from their biological sex; this is usually termed *gender incongruence* (84). The term *gender disorder* has been replaced with the term *gender dysphoria* which describes the distress that an individual might feel as a consequence of having gender incongruence. *Transgender* (often called *trans*) refers to individuals who do not identify themselves as being of their natal gender, whereas *cisgender* (*cis*) people do not experience gender incongruence (85). Readers are also referred to Endocrine Society's 2017 Clinical Practice Guideline and Transgender Health Fact Sheet (84). Estimates of the prevalence of male-to-female transgender individuals among general populations range from 0.5% to 1.3% and estimates for female-to-male transgender individuals range from 0.4% to 1.2% (85). State level population-based surveys indicate that 0.6 % of US adults (25-64 years of age) and 0.7% of adolescents and young adults (13-24 years of age) identify as transgender. Other studies of US high school

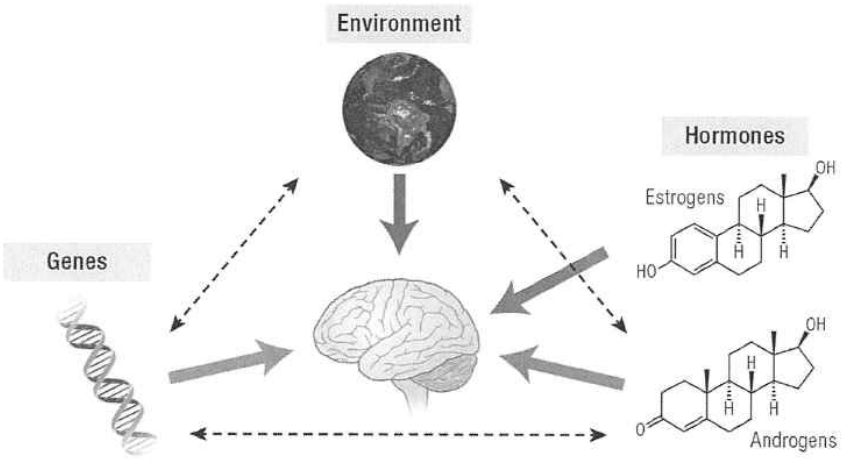


Figure 1. Simplified view of the factors influencing sex differences in the brain. Three broad groups of factors influence the sexually dimorphic brain, as indicated by the broad, colored arrows. 1) Genes and genetic factors that influence the brain include both those on sex chromosomes and autosomes, and include both the DNA itself (represented by the classic double helix) but also chemical modification of DNA (eg, methylation) and modifications of proteins associated with DNA to form chromatin, including histones, and also changes in proteins that bind to DNA. 2) Hormones clearly influence sexual dimorphism in the brain; these are represented by the principal sex steroids, estradiol and testosterone, but also include other steroid and protein hormones (progestins, MIH, oxytocin, prolactin, etc). 3) The environment includes a wide spectrum of influences, including perinatal nutrition and familial support, socioeconomic and demographic factors, intrinsic factors of brain development, age, and gender, and larger environmental factors, such as education, profession, and societal expectations (the "gendered environment"). In addition to each class of factor influencing the brain (bold arrows), the human brain also reciprocally influences each of these groups of factors. Furthermore, each group of factors influences the other, as represented by the dotted arrows. Some examples include: the environment influences genes via epigenomics and genes influence the environment by population sizes and domains; the environment influences hormones by seasonal variations and the actions of xenobiotics, and hormones influence the environment by promoting reproduction and consumption of foodstuffs; genes directly influence hormones by regulating their production and action, and many hormones, including all steroid hormones, regulate gene transcription.

students suggest a prevalence of 1.8% to 2.7% of being gender nonconforming or transgender (86-88). However, several factors may influence reported prevalence of gender dysphoria: (i) small sample sizes; (ii) differences in assessment techniques leading to incomplete ascertainment of gender dysphoric individuals; (iii) unwillingness of some individuals to respond fully and honestly, especially in older studies or studies deriving from locales where gender incongruence is a social taboo; (iv) differences in the subjects ages. *Sexual orientation*, not to be confused with gender identity, refers to the group of persons to whom an individual is sexually attracted; both cisgender and transgender individuals may be hetero-, homo-, or bi-sexual (89).

Although gender is strongly influenced by environmental and cultural forces, it is unknown if the choice to function in society in male, female, or other role(s) is also affected by biological factors (89-91). A general issue is that the association of sex, gender, or sexual orientation with specific brain structures, or with other biological variables, does not establish whether the biological variables are causes or consequences or noncausal correlates of the behavioral characteristics or function of the individuals studied. Three areas of biological difference have been studied fairly extensively: neuroanatomy, genetics, and hormones. Studies have reported differences in the hypothalamic INAH3 nucleus in men vs women and in homosexual vs heterosexual men (92, 93). Although initially controversial, others have confirmed sex differences in INAH3 numbers, not in size or densities, whereas no evidence for sexual dimorphism of any other INAH structures are reported (94). Studies in people with gender dysphoria found that the phenotypes of specific brain structures, such as the bed nucleus of the stria terminalis, of transgender women and transgender men differ from cisgender men and women, with partial, but incomplete sex reversal of sexually dimorphic structures (95). Brain networks involved in one's body perception, (pregenual anterior cingulate cortex, temporo-parietal junction, and fusiform body area) differ in individuals with gender dysphoria compared with cisgender individuals (96-98). Neuroimaging shows that testosterone treatment resulted in functional and structural changes in brain areas associated with self-referential and own body perception (99). Transgender men have thicker medial prefrontal cortex than cis men. Testosterone treatment does not change prefrontal cortex thickness in transgender men, but it has other effects on cortical thickness, connectivity, and fractional anisotropy (99).

Genetics may play a role in gender identity (100): monozygotic twins have 39% concordance for gender dysphoria (101). Attempts to identify specific genes governing gender identity have been plagued by small numbers of subjects and low statistical significance; no specific gene has been reproducibly identified. However, such studies have suggested associations with genes encoding steroidogenic enzymes and sex steroid receptors, and it is generally agreed that androgens play an important but not determinative role. For example, many 46,XX individuals with severe virilizing congenital adrenal hyperplasia (steroid 21-hydroxylase deficiency) are exposed to intrauterine testosterone concentrations typical of those in normal male fetuses and consequently have severely virilized external genitalia; nevertheless, most have a female gender identity, but about 5% to 10% of such individuals have gender dysphoria, an atypical gender identity (89, 102, 103), or atypical sexual orientation and gender behavior (104, 105). Similarly, about half of 46,XY individuals with defects in androgen synthesis who were raised as females revert to a male gender role (106). The biological underpinnings of sexual orientation and gender identity are apparently related but are not the same (107). Thus, there is ample but incomplete evidence for biological substrates—neuroanatomic, genetic, and hormonal—for gender orientation, making this an important area of ongoing research.

Hormonal Versus Sex Chromosome Effects

Sex differences are caused by 3 major factors—sex hormones, genes on sex chromosomes/autosomes, and environment (Fig. 1). To understand disease mechanisms in both sexes and exploit sex differences in protection or exacerbation of diseases, it is important to determine the relative contribution of each of these factors in causing sex differences (17). Many sex differences caused by gonadal hormones have been discovered by measurements of sex steroids and gonadotropins during human development, and in animals by similar measurements or by interventional methods, such as gonadectomy, hormone administration, or the expression of synthetic enzymes or receptors in transgenic mice. Sex steroids play an integral part in many physiological processes (Box 1). Whereas the gonads are the major site of sex steroid synthesis, the adrenals, placenta, brain, and skin can also initiate steroidogenesis, and steroid-modifying enzymes are found elsewhere, especially in liver and fat, permitting synthesis of sex steroid hormones in multiple other sites (50). Thus, animal gonadectomy may provide information about endocrine effects of gonadal steroid hormones but cannot address tissue-specific paracrine effects (Box 2). Moreover, gonadectomy cannot mimic low pre-pubertal levels or physiological conditions in which hormone levels decrease, such as aging or menopause. Manipulations of human gonadal hormones are routinely used in contraception and in the management of sex steroid–dependent cancers (eg, breast, prostate). When

a sex difference is discovered in human disease, and modeled in animals, the investigation of possible hormonal causation of the sex difference is usually the first option considered.

To detect effects of sex chromosomes that cause sex differences, one can compare people who have differences in their sex chromosomes, revealing effects of X or Y chromosome number (108-110). These results strongly suggest direct sex chromosomal contributions to sex differences in cell function. Comparison of brains of XY patients with complete androgen insensitivity (who are phenotypically female), with brains of control XY males and XX females, suggests that cortical thickness and functional connectivity between the limbic regions and the cortex are influenced not only by testosterone actions, but by sex chromosome factors as well (111). However, changes in the sex chromosome ploidy also alter gonadal hormones, so it can be difficult to isolate sex chromosome effects not mediated by gonadal hormone effects. Circulating human embryonic/fetal sex steroid concentrations are poorly characterized, and the tissue concentrations are almost totally unknown. Another approach is to use mice to identify genes on the X or Y chromosome that act outside of the gonads to cause sex differences, and then seek evidence that the orthologous human genes cause human sex differences. Controlled experiments are possible in which XX or XY mice with comparable gonadal hormones can be compared. A frequently used model is the Four Core Genotypes (FCG) model, in which the testis-determining mouse *Sry* gene is deleted from the Y chromosome (creating the Y^- or "Y minus" chromosome) and inserted as a transgene on chromosome 3 (*Sry+*) (Fig. 2 and Box 3) (112). The utility and limitations of these models have been extensively discussed (113, 114).

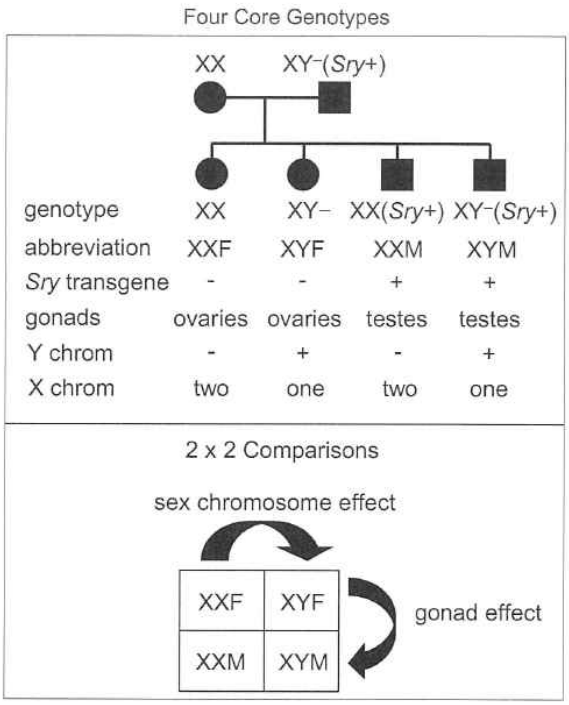


Figure 2. Schematic diagram of the Four Core Genotypes mouse model. The testis-determining gene *Sry* is deleted from the Y chromosome, producing the Y^- chromosome. An *Sry* transgene is inserted onto chromosome 3. Thus, the type of gonad is no longer linked to the sex chromosomes. The model produces XX and XY mice with *Sry* and testes, and XX and XY mice without *Sry*, with ovaries. Sex differences in phenotype can be attributed to an effect of gonadal hormones, comparing mice with ovaries and testes, or to an effect of sex chromosomes, comparing XX and XY mice with the same type of gonad. [Modified with permission from Arnold AP & Chen X. *Front Neuroendocrinol*, 2009; 30(1) © Elsevier Inc. (112)].

Considering Sex and/or Gender as Variables in Health and Disease

Women and men differ in many physiological and psychological variables. It is important to establish the mechanisms causing such differences in health and disease, and to consider sex-related variables in studies of human health and disease. These variables include, but are not limited to, sex- and gender-related factors. The inability to control all variables in human studies means that it may be impossible to determine the relative roles of environment and biology in causing a difference between women and men, when both types of variable can influence the trait. Furthermore, while "gender expression/behavior" can be observed, "gender identity" can only be known by what an individual states. Thus, gender identity, *per se*, cannot be studied in animals. In human studies, it is unethical to selectively manipulate specific biological and environmental variables, and most currently available data derive from studies comparing groups of men with groups of women. It is therefore difficult to disentangle the specific contribution of sex-related genes, hormones, gender-related variables, and other variables that contribute to being female or male. Because sex has long been defined by gonadal type, the list of sex-influencing factors has been primarily associated with gonadal hormones, especially estrogens, progestins, and androgens (121). However, some phenotypic sex differences develop before the gonads differentiate as testes or ovaries (122), so other factors also contribute to sex differences (123) but are seldom considered.

Sex is an essential part of vertebrate biology, but gender is a human phenomenon; sex often influences gender, but gender cannot influence sex. Studies of animal physiology must consider sex as a variable (124), with sex steroids (of both gonadal and nongonadal origins), sex chromosomes, and other factors contributing to sex differences in many physiologic processes. Similarly, studies of human physiology and disease must also consider sex for the same reason (125) and its disorders must

Box 3. Investigating sex chromosome complement versus gonadal hormones in health and disease: the four core genotypes (FCG) model

The FCG model allows for discriminating hormonal vs sex chromosome effects in animals. Gonadal males (XY^-(*Sry+*)), bred to XX gonadal females, produce 4 types of offspring: XY^- and XX mice with the *Sry* transgene and testes, and XY^- and XX gonadal females lacking the *Sry* gene (Fig. 2). Thus, it is possible to compare XX and XY mice with the same type of gonad, in 2 separate comparisons. Differences between XX and XY are attributed to effects of sex chromosome genes acting on nongonadal tissues. To determine if this sex chromosome effect is caused by X or Y genes, a second model is studied, the XY* model (113, 114). This model produces genotypes that are similar to XO, XX, XY, and XXY. An effect of number of X chromosomes is discovered by comparing XO and XX, or XY and XXY. An effect of the Y chromosome genes is discovered by comparing XO and XY, or XX and XXY. These mouse models have been used to demonstrate sex chromosome effects causing sex differences in a wide variety of phenotypes and disease models, including brain and behavioral phenotypes, metabolism, autoimmune, cardiovascular and pulmonary diseases, Alzheimer disease, aging, and cancer (35, 113, 115). These models have facilitated discovery of several disease phenotypes in which the number of X chromosomes contributes to sex differences (116), and a smaller number of sex-biasing effects of Y genes (117). Sex chromosome effects occur in the same disease systems alongside sex-biasing effects of gonadal hormones, such that the 2 effects can synergize to increase the amount of sex difference, or counterbalance each other to reduce a sex difference. Moreover, genes encoded on the Y chromosome can have gene-specific effects, and/or effects that overlap with those of X genes (118). In the cardiovascular system and associated physiological/disease states, sex chromosomes and gonadal hormones can have opposing effects. Estrogens generally protect from cardiac ischemia/reperfusion injury and other cardiovascular diseases, reducing disease in female relative to male mice. However, studies of ischemia/reperfusion injury in gonadectomized FCG mice reveal that the XX sex chromosome complement is associated with worse outcomes, relative to XY (119). In another study, sex chromosome effects in angiotensin II-induced hypertension showed that arterial pressure was greater in gonadectomized XX mice than in gonadectomized XY mice (120). Sex chromosome complement also influences the development of abdominal aortic aneurysms, fat metabolism and adiposity, plasma lipids and lipoprotein levels (particularly HDL-C) (115)).

also consider gender. However, human gender is a spectrum from feminine to gender-neutral to masculine, and also likely includes individuals who do not fit readily on a simple linear continuum (84). Studies addressing the endocrine care of transgender youth during the time of their potential gender transition (84, 89) find that they have a higher prevalence of stress-associated mental health disorders such as depression and anxiety, which can be ameliorated by gender-affirming endocrine treatment (126). It is essential to recognize these sex and gender differences as our health care systems endeavor to develop "individualized medicine."

Despite the fact that biological sex is such a fundamental source of intraspecific variation in anatomy and physiology, much basic and clinical science has tended to focus studies on one sex (typically male). Few studies have done side-by-side testing for sex differences at baseline and in experimental models of human diseases (127-129). Studies in laboratory animals that manipulate biological (eg, genes and hormones) and environmental variables (eg, housing conditions, diet, physical activity, etc) demonstrate that many variables can affect sex-related aspects of an animal's physiology. However, laboratory rodents may show male-female differences caused by different housing conditions, which could be misinterpreted as being caused directly by biological differences without environmental mediation. In studies concerning animal behavior, the sex and gender of the researcher conducting behavioral measures may also influence outcomes (130). Thus, for reproducibility and proper interpretation of the data, at the minimum, it is important to state the precise housing conditions, anesthetics, analgesics (different effects in sexes), doses, surgical manipulations, diet, sex, strain, species, and age of animals used, as well as sex/gender of the researcher(s) performing experiments.

Having laid the foundation for several factors that contribute to sex versus gender, this Statement will use 3 areas of research as examples (not as a literature review) where human and animal sex differences are well known. First, sex differences in specific brain regions of healthy men and women are increasingly being documented along with differences in brain connectomes; these will be discussed in detail in Section II. Second, stress-related pathophysiologies are known to affect twice as many women as men. However, few studies systematically include study designs to ascertain function or mechanisms that may be similar or different between males and females. Hormones and signaling pathways that contribute to sex-specific differences in stress-based pathophysiologies will be discussed in Section III. Similarly, sex differences in manifestation of cardiovascular and renal diseases are well recognized and will be discussed in Section IV.

Section II

Developmental Origins of Sex Differences in Brain Anatomy, Function, and Behavior

Sex differences in the human brain are a topic of intense popular and scientific interest. Several scientific observations motivate the search for sex differences in brain structure

and function. First, the act of sexual reproduction requires that the male and female animals show qualitatively different reproductive behaviors. The stereotyped emergence of these reproductively critical and sexually differentiated behavior reflects biologically programmed (or "innate") sex differences in the organization of those brain circuits that support the motivational and consummatory phases of copulatory behavior (131). Second, the fact that males and females make different biological investments in reproduction—eg, the risks of pregnancy in mammals are borne entirely by the female—sets up sex differences in the behavioral strategies that optimize reproductive fitness (132). Sexual selection based on sex-biased behavioral strategies is predicted to drive the evolution of sex differences in those brain circuits that are responsible for sexually selected behaviors. Third, males and females can show consistent sex biases in broader behavioral domains beyond those that directly relate to reproductive strategies. In our own species for example, there are highly consistent sex differences in the prevalence of physical aggression and violence (both male-biased) (133), as well as extensively documented sex differences in risk for different mental disorders (134).

In this section, we will first describe the main neuroimaging techniques commonly used in comparisons of brain anatomy, connectivity, function, and subnetwork organizations. We then review the key aspects of sex-biased brain anatomy and connectivity that have been revealed by these techniques; sex differences in stimulus-based or task-based functional magnetic resonance imaging (fMRI) studies are not addressed here. Next, we discuss specific disease states that appear to have different outcomes in the 2 sexes due to baseline differences in the "connectome" and animal models used in neuroimaging. Finally, we will address some important caveats and controversies in the field of brain imaging.

Brain Imaging Techniques

Modern neuroimaging methods make it possible to characterize diverse aspects of brain structure, function, and connectivity in vivo. This large toolbox of methods has been used to examine sex differences in brain organization at several levels of analysis. These techniques aim to analyze, map, and visualize regional and inter-regional (connectomic) features of the brain at macroscopic (systems-level) and mesoscopic (neural circuit architecture) levels in order to illuminate brain organization in health and disease (135). Of note, cellular-level details are beyond the resolution of most in vivo brain imaging techniques.

Sex differences in global and regional brain anatomy can be measured in vivo using structural magnetic resonance imaging (sMRI). Several considerations have made sMRI an especially popular technique in the study of brain sex differences in humans. First, sMRI allows a quick and spatially comprehensive screen of the entire brain that can quantify thousands of morphometric properties simultaneously in vivo across a large number of individuals. These characteristics not only facilitate testing for sex differences outside defined regions of interest, but also allow longitudinal measurements that can track the emergence of brain sex differences over development (136, 137). Second, because sMRI considers structure rather than function, it can leverage evolutionary conservation of the basic mammalian brain plan (138), and it is therefore particularly well-suited for cross-species investigation of sex differences in humans and animals. Thus, a critical role for sMRI research in the study of brain sex differences is to screen for brain regions that can then be prioritized for closer analysis using more resource-intensive assays that are typically applied in a regionally selective manner.

Complimenting sMRI, other in vivo neuroimaging techniques such as diffusion MRI (dMRI), resting state functional MRI (rs-fMRI), and fMRI provide unprecedented insights into tissue microstructure and brain connectivity. fMRI maps brain circuitry based on stimulus- or task-based brain functional responses. In contrast, rs-fMRI, by measuring changes in blood flow in the brain generated by signals dependent on blood-oxygen-levels, helps explore the brain's functional organization by providing insights into intrinsic brain activity without requiring participants to be trained in specific tasks, thereby eliminating task performance as a confounder (139, 140). dMRI measures the differential patterns of water diffusivity in biological tissue revealing details of tissue microstructure, especially in white matter (141). Fiber tractography on dMRI enables mapping the fiber architecture of the brain, and subsequently, the network organization of the brain through structural connectomes (142-144). A brain connectome is an extensive map of the white matter structural or functional connections of the brain, created using dMRI or rs-fMRI (145). Modeling efforts, such as the Human Connectome Project, and the use of connectome-based predictive modeling, have provided an integrative, in-depth, and multilevel understanding of the structural and functional connectivity (regions that get coactivated) of the neuronal networks (146, 147).

Sex Differences in Global and Regional Brain Anatomy

It is well established that men have an average total brain volume that is approximately 10% greater than that of women (148, 149). A similar sex difference in average

human brain volume (~8%) appears to be present at birth (150) and is sustained throughout childhood and adolescence (151). The sex differences for total brain volume also hold for the 2 main subdivisions of brain tissue—gray matter and white matter—despite these 2 brain compartments following very different developmental trajectories (151, 152) (Fig. 3).

The robust sex difference in brain volume identified through human sMRI research cannot be fully explained by the fact that brain volume is positively correlated with height (average height is greater in men that in women). Statistical control for body size diminishes, but does not remove, sex differences in total brain volume (149), and boys also show greater average brain volume than girls during early adolescent development, at a time when girls are taller than boys (153). Thus, available literature supports a consistent picture in which there is overlap between the distribution of brain size in men and women, but the mean of this distribution is significantly greater in men than women. The medium effect size of sex on brain volume exists above and beyond sex differences in stature. However, it is important to note that no known functional sex differences associate with the sex difference in overall brain size. Sex differences in overall brain size, and their developmental timing, are both theoretically and methodologically important when considering: (i) whether neuroanatomical sex differences are conserved across species; (ii) whether there are sex differences in regional brain anatomy above and beyond sex differences in overall brain size; and (iii) whether there is concordance between sex differences in brain size and any observed associations between brain size and putative biological causes of sex differences, such as gonadal or sex chromosome status (see below).

The patterning of sex differences in behavior and mental illness risk across the lifespan suggest that sex differences in human brain organization are likely to vary across different brain sub-systems or regions, and potentially also across different developmental periods. Structures in human gray matter compartments mediate neural computation and information processing—in contrast to axon-rich white matter compartments that are primarily involved in connectivity between different brain regions (see "Sex Differences in Brain Network Organization: The Brain Connectome," below). Here, we focus on sMRI studies that have tested for sex differences in regional gray matter volume (regional GMV) after controlling for sex differences in overall brain size. Regional GMV sex differences that survive statistical correction for total brain volume variation are of special interest because they exist beyond global sex differences in brain size. We emphasize GMV rather than other morphometric properties of the brain such as cortical thickness, sulcation, or the shape of subcortical structures (144, 154), because GMV provides a common metric that can be examined across cortical and subcortical structures, with equal applicability to humans and mice. Independent large-scale human sMRI studies in biobanks have identified a reproducible pattern of sex differences in regional GMV using sample sizes that are

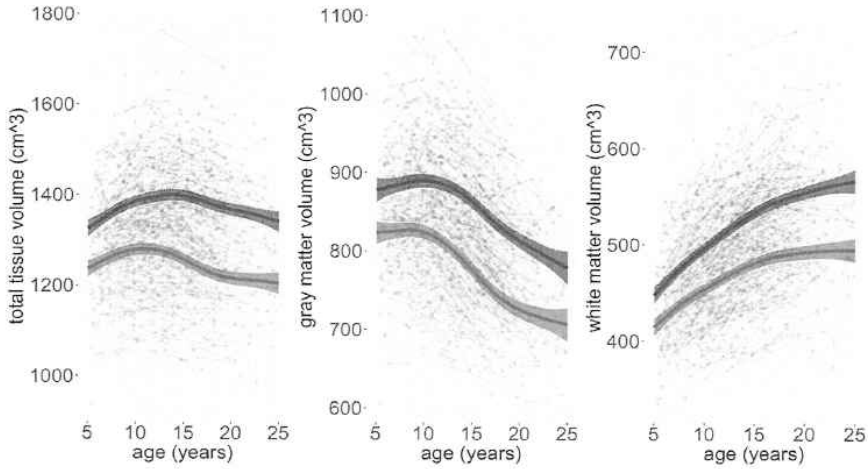


Figure 3. Developmental trajectories for total brain tissue volume, gray matter volume, and white matter volume in men and women over Development. Person-level data are shown for women (red) and men (blue) as points, with lines linking measures from the same person over time. Note the large interindividual variation in volumes within each sex, and the overlap of these distributions, between the sexes. Superimposed on these person-level data are group-level best fit volume trajectories (bold lines with shaded 95% confidence intervals). The developmental window covered is 5 to 25 years of age. For all plots, there are statistically significant sex differences in both trajectory shape (ie, sex differences in the tempo of volume change, $P < 0.00001$), and trajectory "height" (ie, sex differences in absolute volume across ages, $P < 0.00001$). [Adapted with permission from Giedd JN et al. *Neuropsychopharmacology*, 2015; 40 © Springer Nature (153)].

significantly larger than those used in earlier work (148, 149, 155). A structural neuroimaging study involving >2000 individuals demonstrated that higher regional expression of sex-linked genes was coupled with greater GMV in men relative to women (155). These studies, by different laboratories, using different datasets and different techniques for sMRI analysis, find a largely overlapping regional pattern of GMV sex differences after correction for sex differences in total brain volume. These independent replications of regional sex differences in GMV are also in agreement with meta-analytic studies (156). Together, these studies show that, in adulthood, regional GMV is (on average): (i) greater in women than men within superior parietal, dorsolateral frontal, and anterior cingulate cortices; and (ii) greater in men than women within occipital, fusiform, and parahippocampal cortices as well as the amygdala and putamen. Furthermore, while these studies lack temporally resolved developmental maps of male-female differences in regional GMV throughout the brain, there is extensive evidence from focused studies of particular structures that neuroanatomical sex differences can vary dynamically over development, such as observed with amygdala volume and shape (156).

The rapidly expanding body of sMRI research on regional GMV sex differences in the murine brain shows important overlaps and differences with findings from human studies (137, 157). These murine sMRI studies—which are most commonly conducted ex vivo at a spatial resolution of <100 μm throughout the whole brain—have been able to confirm the identification of all classically sexually dimorphic nuclei of male-biased volume from prior histological research, including the bed nucleus of the stria terminalis and medial amygdala (137, 157). These brain regions play a predominant role in modulating social and goal-directed behaviors, pain, and cardiovascular control, all of which are conserved among mammalian species and subject to sexually dimorphic outcomes. By allowing a full-brain screen, murine sMRI has also newly identified a reproducible set of regions with greater GMV in females, including the cerebellar cortex, ventral thalamus, and somatosensory cortex (137, 157). Furthermore, a longitudinal sMRI study in mice found that the set of regions with male-biased GMV can be detected by early postnatal life (with some accentuating over puberty), whereas regions of female-biased GMV in murine adulthood appear to emerge in adolescence (137). To date, there are no studies that formally seek to compare the spatiotemporal patterning of regional GMV sex differences in humans and mice, although existing work already suggests some potential homologies, including foci of greater cerebellar cortex GMV in females vs males by adulthood (137, 148) and the adolescent accentuation of male-biased amygdala volume (158, 159). An important technical challenge in assessing the degree of anatomical homology between regions of sex-biased brain anatomy in humans and mice is that most of the best-established and histologically validated foci of sex-biased brain volume in mice (eg, bed nucleus stria terminalis, medial preoptic nucleus of the hypothalamus) are hard to image in humans due to their small size and intrinsic tissue contrast properties.

Sex Differences in Brain Network Organization: The Brain Connectome

The structural or functional brain network is represented by a "connectome," wherein the structural or functional connectivity between coactivated regions is encoded either through fiber tracts or functional co-activations (160). These connectomes can be studied at the level of subnetworks like visuospatial, auditory, cognitive control, or macro-scale level through global measures of network segregation, integration, and efficiency, to obtain functional associations (161).

A study of 949 individuals (aged 8-22 years; 428 males and 521 females) showed that on average, there are significant differences between the sexes in their structural connectomes (Fig. 4) (162). On average, men had greater within-hemispheric connectivity, as well as enhanced network segregation, whereas between-hemispheric connectivity and network integration predominated in women (Fig. 4A), but these differences were most prominent during adolescence (Fig. 4B-4D). However, an opposite trend was seen for cerebellar connections, which developed differently between human males and females in adolescence and adulthood. The structural connectivity findings were consistent with a behavioral study conducted on the parent cohort (the above-mentioned imaging study was performed on a subset of participants), with women outperforming men on attention, word and face memory, and social cognition tasks, and men performing better on spatial processing and motor and sensorimotor speed tasks (163). An analysis of the Human Connectome Project rs-fMRI data identified age and sex as independent variables that contributed to differences in functional connectivity (164). In brains of men, functional connectivity was more clustered locally in all lobes, except in the cerebellum, whereas the brains of women showed a higher clustering coefficient at the whole-brain level. Thus, brains of men were classified as more segregated and brains of women as more integrated, which agrees with the structural connectivity findings (162). In connectomes, the identification of subnetwork properties (165) can reveal how the complex functional and behavioral repertoire emerges from the simultaneous processes of segregated neuronal clusters and their

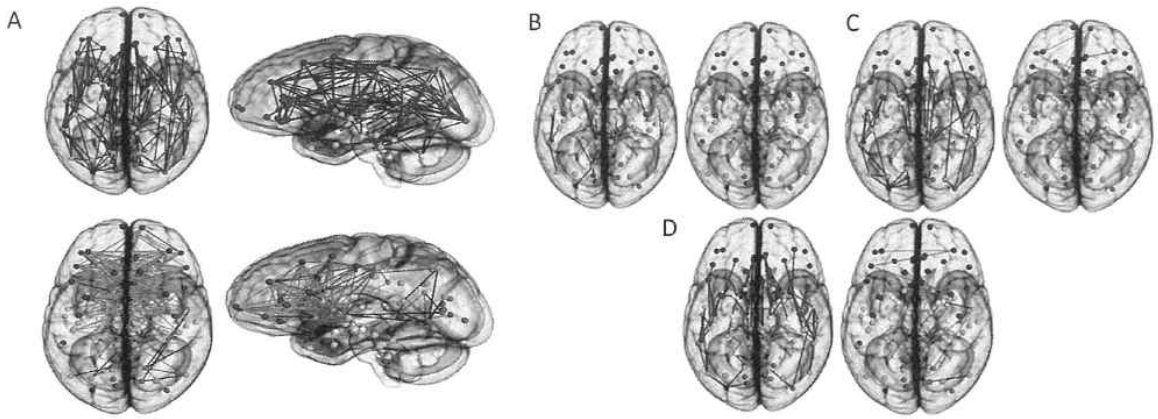


Figure 4. Sex differences in structural connectomes across development. Connectomes representing the white matter structural connectivity in the brain, with nodes indicating the brain regions and edges between the nodes representing the structural connectivity between the nodes. Node colors representing respective brain regions are as follows: dark blue, frontal; cyan, temporal; green, parietal; red, occipital; white, subcortical. The depicted edges shown are those that survived permutation testing at $P = 0.05$. **A,** shows increased intrahemispheric connectivity in men (Upper, in blue) and increased inter-hemispheric connectivity in women (Lower, in orange) on average. **B-D:** Connectivity differences shown in A separated by age groups are shown: **B,** under 13 years, **C,** adolescent (13-18 years), and **D,** young adults (18-22 years). Left image: Men/Boys; Right image: Women/Girls. [Adapted with permission from Ingalhalikar M et al. *Proc Natl Acad Sci U S A,* 2014; 111(2) © National Academy of Sciences (163)].

integration during complicated cognitive tasks (166, 167). Consistent with the behavioral findings on sex differences, men had increased connectivity between motor and sensory (auditory) systems, along with increased connectivity in the fronto-parietal and cingulo-opercular systems that are traditionally associated with complex reasoning and control, whereas women had higher connectivity between reward, memory, and sensory (auditory) systems (163, 168). Better spatial skills in men and improved memory and social cognition skills in women have been reported in behavioral literature (169, 170).

It is important to point out that observed group-level differences in brain structure, function, or connectivity in men and women may reflect the influence of several extraneous factors. For example, in a set of elegant studies, brains of men were imaged to ascertain the contribution of performing complex spatial navigation tasks as part of their daily work on gray matter volume. These studies found that posterior hippocampi of London taxi drivers were significantly larger compared with controls (171), although the work did not address sex differences. Driving a taxi in London before the era of digital maps/navigation systems required extensive training and learning to navigate complex routes before being given a license to operate. In a subsequent study, comparison between London taxi drivers and bus drivers matched and controlled for age, education, intellectual, and stress levels, as well as years of driving experience, showed that taxi drivers had greater GMV in the posterior and less volume in the anterior hippocampi compared with bus drivers (172). Interestingly, years of navigation experience associated with hippocampal volume in taxi drivers alone, but they were significantly worse at acquiring or retrieving novel visuo-spatial information than bus drivers. Importantly, no differences in other GMV, including the caudate nucleus, were found between the taxi and bus drivers; the caudate nucleus is associated with a myriad of cognitive and emotional functions. These studies illustrate brain plasticity and that professional work and years of performing certain tasks can result in brain structural, volume, and connectivity differences that may have little to do with sex or gender per se, but more with training, social environments, and behaviors. In other studies, GMV changes were greater in professional musicians, or after induced training (juggling for 3 months), and in early bilinguals, and white matter volume changes were found in adults learning a second language, irrespective of sex, when reported (173-176). These findings suggest that brain structure retains its plasticity and controlling for factors other than sex or gender are key in interpreting data on structural volumes and associated functions.

The above-mentioned existing datasets did not collect the requisite information on self-report of gender, thereby precluding retrospective analysis of gender in these cases. As identifying correspondence between behavioral scores and the regions that are involved in the manifestation of that behavior remains challenging, analyses of subnetworks pertaining to functional and behavioral domains can help elucidate a brain-behavior correspondence. The detailed description of sex differences in brain organization at the group level, and concerted efforts to specify

the role of sex-biased biological factors in shaping such sex differences, is of fundamental importance (177) and also provides a crucial adjunct for indispensable studies on environmental and wider societal contributions to sex-biased brain development. Such studies should be undertaken jointly using structural and functional connectivity. These studies elucidate the various ways in which sex differences in brain microstructure and connectivity can be investigated.

Sex Differences in Structural and Functional Brain Regions in Obesity

The hypothalamus has long been known as the "center" where peripheral and neural signals converge in the regulation of food intake and energy homeostasis in both sexes. Advances in neuroimaging studies have helped identify activation of several distinct brain regions comprising brain networks in response to eating in men and women. Behavioral and sociocultural factors may play a role in the observed sex differences in ingestive behaviors, appetite, and cravings related to obesity (178). Women report higher prevalence of maladaptive ingestive behaviors such as binge eating, food cravings, and "food addiction," and the lifetime prevalence of disordered eating behaviors are about 3 times higher in women than in men (179, 180). Women also experience episodes of food cravings of greater intensity (181, 182), and greater frequency (183-185), and are less able to suppress food cravings than men (184, 186). Despite the wealth of data indicating that women experience disproportionately higher rates of food cravings, stress eating, and eating disorders than men, the reasons for these differences are incompletely understood (184, 187).

Regulation of food intake entails both homeostatic and nonhomeostatic factors (188). Homeostatic regulation balances energy needs with energy consumption, whereas nonhomeostatic regulation—in particular hedonic regulation and food addiction—involves reward-seeking behaviors that drive humans and animals to consume food beyond their metabolic needs, leading to the development of obesity (189-191). These findings have directed attention toward the extended reward system in obesity-related research, which consists mainly of basal ganglia regions and is involved in dopamine signaling and addiction-like behaviors (192). The extended reward system is composed of 6 interconnected brain networks—salience, central autonomic, basal ganglia, somatosensory, executive control, and emotional regulation (192).

Functional MRI studies have found that, in response to food images, obese individuals show greater activation than normal-weight individuals in regions associated with reward anticipation, dopamine signaling, and addiction-like behaviors (193-196). Greater activity in brain regions of the extended reward network may drive obesity-related behaviors, such as greater responses to food odors and food consumption (197-199). Recent meta-analyses have further supported the role of the brain in disrupting the balance between energy consumption and expenditure. This combination of increased activity in regions associated with reward-driven behaviors and decreased activity in regions moderating top-down control of appetite may lead to consumption of excess calories (188).

Furthermore, sex-specific activations in response to food intake have been observed in cognitive, emotional, and reward-related regions (200-202). For example, obese men had greater activation than obese women in the supplementary motor area, precentral gyrus, fusiform gyrus, and inferior parietal lobule, which are associated with motor control, visuospatial attention, and responding to salient new or alerting stimuli (203). In this same study, obese women showed greater activation than obese men in the caudate and parahippocampal gyrus, regions implicated in reward processing and memory (203). Using graph theory to define the underlying architecture of brain structural connectivity obtained from diffusion tensor imaging, sex differences were observed in the topological measures of centrality (which determine the degree of information flow in specific brain regions) in regions of reward and salience networks in women, and in reward and sensorimotor networks in men (204). Resting state fMRI studies have found sex differences and commonalities in body mass index (BMI)-related connectivity associated with specific defined regions of interest in the reward network (205). For example, women had increased associations between BMI and increased connectivity in the in right globus pallidus and bilateral putamen. In men, BMI was associated with increased connectivity in the medial frontal cortex. A study of sex differences in response to visual and auditory food cues found that women experience greater activation in lateral and dorsolateral prefrontal and parietal cortical regions involved in cognitive planning and executive guidance and evaluation of behavior, compared with men (202). When viewed together, these studies highlight the importance of investigating sex differences in obesity-related alterations in the core and extended reward networks.

Although many single-sex studies of fMRI and obesity have been published, with the majority having all-female subjects, few studies have specifically investigated sex differences in brain function and structure in obesity. Despite the literature supporting sex differences in the brain, including in regions implicated in reward behaviors and energy homeostasis, few comprehensive reviews of sexually dimorphic brain signatures related to obesity have

been performed. A recent meta-analysis using an activation likelihood estimation approach to evaluate comparisons in functional responses to stimuli by obesity and by sex revealed differential sex- and BMI-related activations in reward anticipation and response, in shaping food-related memories, and in generating top-down control of appetitive processes. Together, these findings have important implications for sex-specific obesity treatments.

Models to Study Sex Differences in Normal Brain Structure and During Pathophysiology

Studies of sex differences offer important considerations for personalized medicine. The prevalence, clinical presentation, and symptomatic progression of many neurological and psychiatric disorders are remarkably different between the sexes. In addition to common X-lined mental retardation syndromes, men have a greater prevalence of neuropsychiatric disorders such as autism, attention-deficit/hyperactivity disorder (ADHD), and Tourette syndrome (206), whereas women have a greater prevalence of mood and eating disorders (207, 208). From the perspective of developmental disorders, the differences in the developmental trajectories of the sexes perhaps represent different vulnerabilities of maturing brain circuitry, leading to differences in symptoms, onset, and severity of neurological disorders. There are also sex differences in the risk factors, average age of onset, and prevalence of late-life dementias, as well as cerebrovascular disease (209). Additionally, in traumatic brain injuries, where the network organization of the brain is affected by the injury, such as the corpus callosum region, sex differences in inter-hemispheric connectivity and brain subnetworks may influence the impact of injury, and hence subsequent recovery. Thus, sex differences in brain connections are crucial to identify, as they may elucidate mechanisms in disease risk and potential treatment and recovery (210).

Most models of sex-biased mammalian brain development are based on experimental data from rodents (now largely from mice, but previously also from guinea pigs and rats). One of the most systematic dissociations of gonadal and chromosomal contributions to sex-biased anatomical brain organization in mammals is provided by a recent sMRI study of adult mice from the FCG model (112, 211). By combining sMRI with behavioral assays, these studies determined the contribution of sex chromosomes and gonads to adult mouse brain structure and function (211). This study revealed: (i) an effect of sex chromosomes on regional GMV in the cerebellar cortex and olfactory bulb; and (ii) an effect of gonads on regional GMV in the parietotemporal cortex and the bed nucleus of the stria terminalis. Some of these effects overlapped with regions of normal sex differences in murine GMV (eg, cerebellar cortex and bed nucleus of the stria terminalis), and some brain regions were anatomically sensitive to both effects (basal forebrain and periaqueductal gray matter). Sex-chromosome effects on regional gray matter anatomy have also been reported by complementary sets of sMRI studies in both mice and humans that compare groups of euploid individuals with groups carrying X-chromosome aneuploidy (157, 212). Finally, in both mice (137) and humans (155), the spatial patterning of sex differences in regional GMV in adulthood appears to be preferentially aligned with the spatial patterning of sex-chromosome gene expression—which points toward a potential role of sex-linked genes in the establishment of maintenance of regional GMV sex differences. These studies emphasize the need for integrative models that view biological contribution to sex-biased brain development as a developmental dance of coordinated influences from both gonads and sex chromosomes.

Caveats and Critiques Relating to Neuroimaging of Brain Sex Differences

While several sMRI studies apparently establish that there are highly reproducible male-female differences in regional gray matter volume after controlling for variation in total brain size in humans, this conclusion should be considered in the light of several important caveats and critiques to avoid misinterpretation. First, all sMRI phenotypes that show reproducible and statistically significant sex differences also show a considerable overlap between men and women. This overlap is illustrated by total brain volume: total brain volume averages 10% greater in men than women, but many women have a total brain volume above the 30th centile for male brain volume, and many men have a total brain volume below the 30th centile for female brain volume (149). Sex differences in brain structure and organization are present across the lifespan and vary based on age, so inferences should be drawn cautiously. Thus, while total brain size shows a robust mean difference between men and women, an individual's total brain volume is a weak predictor of biological sex. These 2 facts arise because biological sex is only one source of variation in brain size (149), and other factors/variables that influence total brain size are unknown and/or hard to model statistically (Fig. 1). By extension, because sources of anatomical variation can differ between brain regions—the same individual can have GMV values that appear to be "sex-typical" in one region, but "sex-atypical" in another (when typical and atypical are defined by an individual's percentile position relative to the distribution of population-level trait variation in each sex) (213). This interpretation offers one

potential explanation for the observation that an individual brain can show varying degrees of GMV "sex-typicality" in different brain regions (relative to the population distribution). Alternative explanations have been proposed, including regional variations in programs of sex-biased development such that one individual's brain may be considered a "mosaic" of male and female parts regardless of their chromosomal and/or gonadal sex (213).

Second, although sex differences in regional GMV are highly reproducible in humans and mice, these mesoanatomical sex differences *cannot* be assumed to correlate with behavioral sex differences. The functional relevance of neuroanatomical sex differences is hard to establish experimentally in humans, but correlations between anatomical and behavioral sex differences could be modeled in humans using several feasible study designs. To date, however, very few studies have directly tested for such structure-function correlations in humans (161), and this is an important priority area for future research. Several other challenges will need to be addressed in future work for any given sex-biased sMRI phenotype, including which aspects of behavior to measure and how to consider properly all possible configurations of brain-behavior association in 2 groups (eg, varying intercepts and/or regression slopes across groups). Moreover, some sex-biased sMRI phenotypes, such as trajectories of anatomical change, can only be estimated from group-level data, which complicates comparisons with interindividual variation in behavior. More fundamentally, however, regional GMV sex differences may be useful for understanding the brain basis for sex-biased behavior without GMV variation itself being the behaviorally relevant marker. For example, sex differences in mean regional GMV may help to define brain circuits that subserve sex-biased behaviors through their molecular, cellular, or connectivity features rather than through their volume *per se*. It is also important to entertain the possibility that sex differences in the anatomical organization of a given brain system may actually serve to equilibrate function between the sexes despite each sex having a categorically different genetic starting point.

Third, in addition to the functional considerations above, full understanding of a given sex bias in regional brain anatomy requires a mechanistic account that can link observed anatomical sex differences back to specific genetic and/or environmental factors that differ between men and women. It is usually impossible to disentangle biological sex differences from those which could be the result of environmental influences during development, differences in gender, and in sexual orientation (Fig. 1). Strict causal tests for mechanistic models of sex-biased brain development are very hard to achieve in humans, although several informative approaches have been pursued including: (i) modeling sMRI data using normative variation in hypothalamic-pituitary-gonadal axis maturation or function (214); (ii) applying sMRI methods to cohorts undergoing gender-reassignment (215); and (iii) studying how sMRI features differ between typically developing groups and those affected by medical disorders involving the sex chromosomes (eg, sex chromosome aneuploidies) or sex steroids (eg, androgen insensitivity, congenital adrenal hyperplasia) (215, 216). However, the opportunistic and correlational nature of these approaches places considerable limits on the inferential power of mechanistic studies of human sex-biased brain development. Moreover, as challenging as it is to study chromosomal or gonadal factors in humans, it is even harder to address empirically the many plausible hypotheses about the potential for experiential and societal influences to differentially shape brain development in both sexes (121) or genders.

Section III

Sex Differences in Molecular Mechanisms Underlying Brain-Gut Disorders

The brain and the gut communicate with each other in a bidirectional way through parallel and interacting channels, involving immune, endocrine, and neural signaling mechanisms (217). The brain is able to modulate gut permeability, motility, intestinal transit, and microbial function via the autonomic nervous system (217), and the gut in turn sends signals to the brain to modulate behavior, in rodents (218). This brain-gut communication is especially critical in mediating stress responses and in stress-based disorders. In psychiatric and other neurological diseases, there are notable sex differences that point to different underlying neurobiological mechanisms in men vs women (219-221). Despite their clear documentation, these sex differences have largely been ignored, in order to develop broadly applicable pharmacotherapies that come at a considerable cost, especially for women's health (222, 223). Sex biases in psychiatric risk are particularly instructive as they are developmentally patterned in a manner that is highly reproducible across different cultural settings and historical epochs: early-onset neurodevelopmental and gut disorders are more prevalent in boys than girls, while the opposite sex-bias is seen for adolescent-emergent mood disorders (134, 224). Brain-gut disorders are more prevalent in women than men, but this may be due to underreporting by men due to social stigma associated with several of these

disorders. The etiologies and risk factors for several brain-gut disorders differ between the sexes, yet study designs include predominantly male sex. In this section, we discuss the possibilities that shared and distinct mechanisms operate in males and females resulting in similar as well as distinct manifestation of symptoms for a given disease/disorder.

Sex-Related Differences in Obesity

Although prevalence rates for obesity are at unprecedented levels in all ages (225) and are almost equal in men and women (except when stratified by race or ethnicity) (226), recent surveys indicate an increase in the incidence of obesity in adults and sex differences in the associations between weight, physical health, and psychosocial functions (227, 228). Sex differences in body fat distribution have also been observed (178, 229), with women showing an increased propensity to gain total body fat, especially subcutaneous abdominal fat, whereas men tend to have more visceral adipose fat (230), which is associated with higher risks of type 2 diabetes, hypertension, dyslipidemia, and cardiovascular disease (231). Most clinical trials do not report sex differences related to health outcomes or treatment responses, but a few existing reports suggest women are less likely to complete treatment, tend to lose less weight than men, have a greater number of unsuccessful attempts to maintain weight loss resulting in the well-known "yoyo" diet phenomenon, and have limited responses to pharmacological treatments (225). Obesity-related studies in humans and rodents have expanded in scope to not only focus on structural and functional brain differences between obese and lean male and females, but also include investigations into the bidirectional signaling associated with the brain-gut microbiome axis (232, 233). In obese individuals, changes in the relative abundance and gut microbial diversity have been linked to changes in metabolism, insulin resistance, inflammation, and fat deposition (234). The importance of the intestinal microbiome to human health has been of interest over the past few decades, with multiple studies now linking the microbiome to energy homeostasis, immune function, and development of obesity and metabolic syndrome (235-237), even though few studies have addressed causality.

Not only does the brain-gut axis demonstrate changes in obese individuals, but evidence also highlights differences in the microbiota based on sex hormones (238). More recently, the effect of sex hormones on the composition of the gut microbiota has been explored, with differences seen in the microbiota between men and women during various stages of human development and maturation (238). These sexually dimorphic microbiome signatures are likely to contribute to differences in susceptibility to autoimmune and metabolic diseases between the sexes. Studies performed in immunocompromised mouse models have shown delayed onset and lessened severity of type 1 diabetes in female mice who receive male microbiota transplants; testosterone activity and androgen receptor signaling was essential for this protection (239, 240).

These sex-specific differences in the microbial communities persist throughout adult development, with murine models demonstrating the role of testosterone in orchestrating these divergences in host selection of microbial communities (240). In rodents, males exhibit lower microbiome variability relative to females, likely due to the pulsatile nature of estrogens (240). Human studies comparing the microbiome of twins also revealed more divergences in microbial composition in opposite-sex versus same-sex twins (241). When the cecal contents from adult male mice is transferred into female mice, metabolomic profile changes and masculinization of the hormonal profile results, suggesting the gut microbiota's influence on sex-specific metabolic and behavioral phenotypes (239, 242).

Circulating estrogens in the body are metabolized by the liver and undergo methylation, hydroxylation, and conjugation reactions to produce metabolites that affect host metabolism (243). Certain metabolites are excreted through the bile and are further processed by microbial enzymes in the distal small and large intestine. Certain microbial species secrete beta-glucuronidase, an enzyme that deconjugates biliary estrogen metabolites and allows for its reabsorption into the bloodstream to act on distal sites through binding of estrogen receptors (244). Dysbiosis and decreased microbial diversity result in decreased production of absorbable estrogen metabolites. This mechanism has been implicated in pathologies associated with low circulating estrogens, such as obesity, metabolic syndrome, cardiovascular disease, and cognitive decline in women (245, 246); however, estrogen replacement therapy does not reverse these conditions (247). Growth hormone similarly contributes to sexually dimorphic responses in the above-mentioned diseases (248). In addition, estrogens modulate inflammatory pathways driving disease processes such as nonalcoholic fatty liver disease (NAFLD) and type 2 diabetes (249, 250). More specifically, estrogens regulate adipokines and lipopolysaccharides, which respectively are adipocyte-derived hormones and endotoxins that have been associated with type 2 diabetes (251). Adipokines play a role in metabolic homeostasis as well as in mediating the beneficial and detrimental effects of inflammation (252). The androgen- and estrogen-dependent regulation of adipokines, including leptin, resistin, adiponectin, and visfatin, provides a possible mechanistic link between metabolic disorders (obesity,

atherosclerosis, insulin resistance) and autoimmune dysfunction. The estrogen-microbiome axis can provide a potential avenue for a sex-specific approach to combating metabolic disorders and highlights the bidirectional interaction of estrogens and microbial communities in the pathogenesis of disease processes.

Although the exact signaling mechanisms underlying the communication within the brain-gut-microbiome axis remain incompletely understood, tryptophan metabolites have been implicated as important signaling molecules (253). The most extensively studied tryptophan metabolite is serotonin (5-HT), a molecule with diverse roles in both the gastrointestinal tract (ie, peristalsis, secretion, and absorption) and the central nervous system (ie, mood, pain modulation, behavior, sleep, and ingestive and cognitive functions) (254). Tryptophan also acts as a precursor to the kynurenine (KYN) family of molecules (255). In obesity, the KYN pathway is preferentially activated and may contribute to immune-mediated inflammation, which may drive inflammation-associated changes to the extended reward network described in previous brain studies, particularly changes involving the amygdala and lateral orbitofrontal cortex (256-259). KYN may also modulate signaling within the brain-gut-microbiome axis through downstream neuroactive metabolites, such as kynurenic acid and quinolinic acid, functioning as N-methyl-D-aspartate (NMDA) antagonists and NMDA excitotoxins, respectively (260). Sex differences have been reported in these metabolite products in obese individuals, with lower tryptophan levels but elevated KYN and KYN/tryptophan ratios in women with high BMI compared to men with high BMI (256, 261, 262).

Sex Differences in Stress-Based (Patho) Physiologies

Epidemiological data reveal that the majority of psychiatric disorders occur at different rates in men and women. For example, men are more likely to suffer from attention-deficit/hyperactivity disorder (ADHD), whereas women are more likely to suffer from major depression and posttraumatic stress disorder (PTSD) (219, 263-265). Even when the rates of disorders are similar, their presentations can differ. Schizophrenia, for example, is only slightly more common in men than women, but men develop schizophrenia at an earlier age and present with more negative symptoms, such as social withdrawal and lack of motivation. (224). In the case of bipolar disorder, rates are similar between the sexes, but women more often have more rapid cycling and mixed episodes and they report higher comorbidity with eating disorders and PTSD, whereas men report higher comorbidity with alcoholism (266). Not only does the risk and presentation of psychiatric disorders vary between men and women, but there are differences in treatment responses. For example, the efficacy of antidepressants differs between the sexes: men respond better to tricyclic antidepressants, whereas women respond better to selective serotonin reuptake inhibitors (267, 268). These findings implicate neurobiological sex differences in contributing to disease. In support of this idea, recent studies using animal models are beginning to uncover molecular processes that can bias males and females toward different pathology. Findings from some of these basic research studies will be highlighted here as examples of how including sex as a biological variable can inform our understanding of the etiology of stress-based disorders, as well as guide the development of better treatments.

While there are sex differences in rodent studies in the structure and the size of certain brain regions that can contribute to sex differences in behavior (211), imaging studies that focused on sex differences in cortical thickness and gyration suggest a role for these brain regions in humans as well. In adolescent girls, cortical thinning in the right temporal regions, the left temporoparietal junction and the left orbitofrontal cortex is faster than in boys (154). In contrast, changes in cortical folding were only found in one cluster of the right prefrontal region, suggesting that the mechanisms underlying changes in cortical thickness and gyrification in adolescents are distinct. Sexual dimorphism in the developmental course of the cortical maturation, which coincides with the onset of puberty, might explain sex differences in the age of onset and clinical presentation of many psychiatric disorders (154). Recent evidence has revealed that molecular sex differences in the brain are more widespread than initially thought and such seemingly small-scale differences can have a large impact on physiology and behavior (269). Neurons typically communicate with each other via neurotransmitters and neuropeptides, which are released from a presynaptic neuron and travel across a synapse to bind to receptors on the postsynaptic neuron to exert downstream cellular effects. There are sex differences in production and release of many neurotransmitters and neuropeptides that can result in behavioral changes. In other instances, sex differences in these systems are compensatory, leading to similar behavior endpoints via different mechanisms. For example, both male and female juvenile rats play, but the release of the inhibitory neurotransmitter gamma-aminobutyric acid (GABA) into the lateral septum mediates juvenile play only in female rats (270). There are also sex differences in receptors that can influence how these neurochemicals affect their downstream targets. For instance, dopamine 1 (D1) receptors, which belong to the family of G protein-coupled receptors (GPCRs), in the nucleus accumbens, are necessary for social

withdrawal in female but not male California mice (271). The function of GPCRs is often complex and they can induce different downstream effects depending on their conformation and location. Sex differences can occur at each level of receptor function, in some cases altering physiology differently in male vs female rodents. Sex differences in GPCR signaling are particularly important to consider, especially given that GPCRs are the most studied drug target family for a myriad of indications; in fact, 34% of all US Food and Drug Administration (FDA)-approved drugs are targets of GPCRs (272). As an example of the myriad of sex differences that can be mediated by receptors, we will use the corticotropin-releasing factor 1 and 2 (CRF_1 and CRF_2, respectively) receptors that facilitate responses to stress, exhibit sexually dimorphic expression pattern, are modulated by both estrogens and androgens, and have been relatively well characterized in both sexes (273, 274).

Upon perception of stress or perturbation of homeostasis, CRF is synthesized in the paraventricular nucleus and released from the median eminence of the hypothalamus into the pituitary portal circulation, which in turn stimulates the synthesis and secretion of adrenocorticotrophic hormone (ACTH) from the anterior pituitary into the general circulation. ACTH acts on the adrenal cortex to stimulate the synthesis and release of glucocorticoids and other steroids. This activation of the HPA axis in the classic "flight or fight" response by the CRF system is present in all mammals. The mammalian CRF family comprises 4 agonists, CRF and 3 urocortins (UCN1-3); and 2 known class B GPCRs, CRF_1 and CRF_2. While CRF_1 and CRF_2 share ~68% identity at the amino acid level (275), they perform distinct functions; CRF binding to CRF_1 initiates stress responses by activating the HPA axis, whereas UCN1-3 binding to CRF_2 brings systems back to homeostasis (274). Not surprisingly, perturbations in the components of the CRF family impact several organs and lead to brain-gut disorders, type 2 diabetes, metabolic syndrome, cardiovascular, and reproductive diseases, among others (274). There are sex differences in CRF's endocrine effects. In female rats, higher levels of CRF mRNA in the paraventricular nucleus are reported that associate with the estrous cycle (276, 277) and are reviewed elsewhere (274). Perhaps as a compensatory response, CRF binding protein, an endogenous protein that sequesters CRF thus preventing its bioavailability, is expressed at higher levels in the pituitary of female compared with male mice (278). In humans, there is evidence for increased CRF receptor sensitivity at the level of the pituitary of women relative to men, because peripherally administered CRF, which acts at the pituitary, increases ACTH to a greater degree in women (279).

During stress, CRF is also released centrally into many brain regions, where its neuromodulatory effects coordinate cognitive and behavioral changes to promote stress coping (280). There are sex differences in the way these brain regions respond to CRF that are largely due to sex differences in CRF receptor signaling (274). For example, there is greater CRF_1 receptor binding in the basolateral amygdala in female rats (281). In contrast, binding of the CRF_2 receptor subtype, which is involved in stress recovery, is greater in the central nucleus of the amygdala in male rats (281). It is unknown precisely how these sex differences affect behavior, but given that the amygdala is critically involved in fear, it is likely that these receptor sex differences differently alter fear processing in males and females. In the brain, CRF_2 is most abundant in the bed nucleus of the stria terminalis, a region that regulates sexual behavior and stress-related functions (282, 283). Promoters in genes for CRF_1 and CRF_2 receptors harbor estrogen and androgen responsive elements and show tissue-specific modulation by sex hormones (284, 285). The sexually dimorphic expression pattern of these receptors at normal physiological states and during stress or disease pathology are summarized in a recent review (274).

Sex differences in CRF_1 receptor signaling have been identified in the noradrenergic-containing nucleus of the locus coeruleus (LC) and these differences have important implications for understanding disease vulnerability (273). The LC-noradrenergic system regulates levels of arousal such that higher levels of norepinephrin are associated with greater levels of arousal (286-289). Stressor exposure causes CRF to be released into the LC, which speeds up LC neuronal firing, increasing norepinephrin release (290, 291). Activation of this system during an acute or moderate stressor is thought to be adaptive, because it is important to be alert during a stressful event. However, if this system is activated inappropriately or persistently it can lead to hyperarousal that contributes to agitation, restlessness, impaired concentration, and sleep disturbance. Hyperarousal is a key feature of PTSD and reported in a subset of depressed patients (292, 293). Similar sex differences in spatiotemporal expression of CRF_2 and its ligands are found in humans with gut disorders, where they could contribute to differences between males and females in vulnerability to brain-gut disorders (127, 294).

There are sex differences in CRF_1 receptor signaling in the LC that increase female sensitivity to CRF. In the LC, CRF receptors primarily couple to Gs to initiate signaling through the cyclic adenosine monophosphate (cAMP)-protein kinase A (PKA) signaling pathway (295-297). Sex differences in CRF_1-induced cAMP-PKA signaling are linked to greater coupling of the CRF_1 receptor to Gs in females compared to males (298). This sex difference in

coupling of Gs may indicate that the CRF_1 receptor has a different conformation or binding partner in females vs. males, permitting different proteins to preferentially bind in each sex. Further support for this idea comes from studies demonstrating that, in male rats, acute swim stress increases the binding of a different protein, β-arrestin2, to the CRF_1 receptor, and this effect is not observed in female rats (298). The increased β-arrestin2 in male rats likely contributes to the greater CRF_1 receptor internalization in stressed males (298). When taken together, these findings suggest that CRF_1 receptors preferentially signal through different pathways in males (small GTPases) and females (cAMP-PKA) (299). This difference in signaling could alter physiology and disease risk. In fact, sex differences in CRF_1 receptor signaling in cortex were linked to increased Alzheimer-related pathology, including increased tau phosphorylation and amyloid β signaling in female compared with male mice (300). Few studies investigate sex differences in GPCR signaling, but it is likely that sex differences in GPCRs are also found in receptors other than CRF and that these differences could confer vulnerability and resilience to many diseases.

In human studies, single nucleotide polymorphisms in the CRF receptor gene (*CRHR2*) are associated with negative emotions in patients with irritable bowel syndrome (IBS) (301). Immune cells secrete CRF_2 in extracellular vesicles that circulate in the plasma and associate negatively with disease severity scores in IBS-diarrhea patients (294). Single nucleotide polymorphisms in *CRHR2* are also associated with lifetime PTSD in women (302) and with type 2 diabetes (303). The prevalence of type 2 diabetes and insulin resistance is greater in men (304). Epidemiological studies have shown that men with high levels of self-reported perceived stress have a 1.4 higher odds ratio of developing type 2 diabetes during a 10-year follow-up period and are at 2-fold higher risk of developing diabetes than women with similar levels of reported stress (305). In agreement with human data, male mice lacking functional stress receptors (*Crhr2*$^{-/-}$) and haploinsufficient (*Crhr2*$^{+/}$) mice have worse glucose and insulin tolerance, microvesicular hepatic steatosis, and dyslipidemia than female *Crhr2*$^{-/-}$ or C57BL/6 male and female mice in a high-fat diet–induced model of diabetes (129). Female *Crhr2*$^{-/-}$ mice had significantly greater brown adipose fat mass on high-fat diet than C57BL/6 female or male mice of either genotype, suggesting greater thermogenic responses that might be protective. However, the mouse study did not address whether steroid hormones contributed to changes in adipose mass or function. Thermogenesis in brown adipose tissue in humans in response to a meal or cold stress suggests that women have greater thermogenic responses than men and that these responses correlate positively with progesterone levels, but negatively with cortisol levels (306). Thus, analyzing data from both sexes provides insights into sex-specific mechanisms that regulate physiological processes in both sexes.

In colonic tissues of pediatric patients with Crohn's disease, subcellular localization of CRF_2 differs between boys and girls (127). Furthermore, lack of CRF_2 revealed several sex-specific signaling pathways and differential degree of inflammatory responses in male and female mice (127). Treatment with UCN1, a high-affinity agonist for both CRF receptors, rescued *Crhr2*$^{-/-}$ male mice from colitis-induced mortality, whereas UCN1 treatment increased mortality in *Crhr2*$^{-/-}$ female mice (127). Both diabetes and Crohn's disease show sex differences in disease prevalence and outcomes, yet most animal studies use male sex to delineate mechanisms. Analysis of the data by segregating the 2 sexes can reveal significant insights into distinct and shared mechanisms and factors that exist at baseline and during disease. For example, sex differences exist in the etiology of pancreatitis: alcohol and tobacco predominate in men, whereas idiopathic and obstructive etiologies predominate in women (307), yet to date only a few studies have used both sexes to study mechanisms involved in pancreatitis. While both males and females develop pancreatitis in animal models, when administered identical doses of the pancreatic stressor caerulein, C57BL6 female mice show less severe pancreatitis and histological damage than male mice (128). Lack of CRF_2 rendered female mice more susceptible to caerulein-induced pancreatitis compared with male *Crhr2*$^{-/-}$ mice (128), with both male and female *Crhr2*$^{-/-}$ mice exhibiting similar levels of total histological damage (128). Detailed analysis of components contributing to histopathological damage showed that female C57BL6J mice have less necrosis, zymogen granules, and vacuolization than male mice with pancreatitis, but they have similar levels of edema and neutrophil infiltration as male mice (128). This data segregation allowed isolation of factors that differentially contribute to histological damage, which otherwise would be lost, if grouped together in this analysis. Taken together, these data support a role for the CRF receptors, product of an autosomal gene and regulated by steroid hormones to bring about sex-specific cellular signaling and function.

Sex Differences in Pharmacotherapy of Stress-Based Diseases

Sex differences in GPCR signaling are also relevant for pharmacology. Biased ligands can shift signaling toward

β-arrestin pathways and away from G-protein-mediated pathways based on how they bind to the GPCR (308). These biased ligands are being designed with the hope of providing more targeted therapies with fewer side effects (308, 309). Understanding sex differences in signaling and how such differences contribute to changes in physiology can inform the development of these biased ligands. For example, a CRF_1 receptor ligand that biases signaling through β-arrestin pathways may be useful for treating hyperarousal symptoms or reducing the progression of Alzheimer disease, especially in women. An idea for such a compound would never have come about if women were excluded from preclinical and clinical studies on CRF_1 receptor function.

The idea of using CRF_1 antagonists to treat depression, PTSD, and irritable bowel syndrome has been around for decades, but these compounds were ineffective in several clinical trials (222, 310). Sex differences in CRF_1 and CRF_2 receptor signaling may also explain the failure of different selective CRF_1 antagonists as treatments for these disorders. While there are likely many reasons for their failure, critical ones could be sex differences in their target, association of CRF receptors with different binding partners in female versus male cells, or heteromerization of CRF receptors (311-313), all of which can result in altered signaling. The consistent efficacy of CRF_1 antagonists in reducing anxiety-like and depressive-like behavior in rodents and nonhuman primates was established in studies primarily conducted in male animals (222, 314-317). In a study in which females were included, local blockade of CRF_1 receptors in the dorsal raphe with an antagonist reduced anxiety in male but not female mice, highlighting sex differences in efficacy (318). Yet these compounds developed primarily in male rodents were tested in clinical trials with participants of both sexes or only in women. Notably the only CRF_1 antagonist study that had success in reducing depressive symptoms, NBI-34041, was conducted only in men (222, 319). The approach of developing compounds in male animal models is not unique to CRF_1 antagonists and has been common practice (222). Collectively, these studies suggest that a failure of certain therapeutics may result from ignoring sex differences in their targets. Sex differences in targets are not well known because most preclinical studies use only male rodents (320, 321). Excluding females in the drug development stage particularly impacts women's health. Indeed, it is likely that some compounds deemed ineffective in male rodents would work in females, yet such compounds never would have a chance to make it to market, because of testing exclusively in male subjects. Moreover, the fact that most drugs are designed using males also likely contributes to the higher rates of adverse drug reactions in women compared to men (322).

Including both sexes in mechanistic studies is critical for developing drugs that work efficaciously in both sexes (see Box 4). Latent sex differences can also impact drug development: a compound targeting a mechanism in men may not work in women. As the field moves forward, we may find that sex-specific therapeutics based on understanding latent sex differences are required to truly improve patient outcomes. In sum, there are observable sex differences in behavior that extend beyond reproductive function. Molecular sex differences in several organs, such as the gut and the central nervous system, play a key role in driving these functional and behavioral differences. Moreover, even when function and behavior are consistent between the sexes, the underlying processes can differ. Thus, including both sexes in preclinical molecular studies guiding drug development is key for improving the health of men and women.

Section IV

Sex Differences in the Cardiovascular-Renal System

Cardiovascular disease (CVD) is the major cause of premature death in both sexes worldwide, although women generally develop CVD 10 years later than men (328). In 2016, ~18 million people died from CVD, representing ~30% of all deaths worldwide (329). There are marked sex differences in CVD and renal disease. For example, women are protected from heart disease during the reproductive years but are more likely to die in the first year following a cardiovascular event than males (330). Most heart conditions, including myocardial infarction, Takotsubo syndrome, and cardiac arrythmia, exhibit sex differences in symptoms and severity (331). Chronic kidney disease (CKD) is more prevalent in women but, once established, progresses more rapidly in men (332). However, this female advantage is lost after menopause. These sex differences in cardiovascular and renal disease have long been overlooked and underappreciated. The clinical presentation, the response to pharmacotherapies, standard care practices, and the underlying pathophysiological mechanisms differ in women compared to men. Furthermore, lack of understanding of sex differences in mechanisms underpinning cardiovascular and renal disease has led to poorer outcomes in women than in men. A major problem is that mechanistic preclinical studies in animal models have largely been conducted in males (333). Yet, it has become increasingly clear that sex differences

Box 4. Sex differences in pharmacokinetics and pharmacodynamics of drugs

Thalidomide, a sedative that was prescribed to many pregnant women to relieve pregnancy-associated nausea, was first sold in Germany (without a prescription) in 1957; it had been tested in animals and in men, but not in women. It was soon noted to cause multiple birth defects, most notably phocomelia (arrested limb development) and postnatal deaths. Fortunately, it was never approved in the United States, but thousands of children were affected around the world. In 1962, the US Congress passed the Kefauver-Harris Drug Amendments Act requiring manufacturers to prove a drug is both safe and effective (323). Consequently, the US Food and Drug Administration (FDA) recommended against drug testing on women, particularly those of child-bearing age, until the early 1990s. To date, most treatment guidelines are based on results from clinical trials conducted on middle-aged men. Dosage, pharmacokinetics, and pharmacodynamics data for women (and children) are lacking for most drugs. Activities of cytochrome P450 (CYP) enzymes show significant sex differences in drug metabolism in Phase I clinical trials (324). Gastric enzymes involved in oxidative degradation such as alcohol and aldehyde dehydrogenases are significantly more active in men than in women resulting in higher bioavailability of ethanol in women versus men. In Phase II trials, glucuronidating enzymes and some efflux transporters have been shown to be more active in men then in women. Together with estrogens and androgen that alter transmembrane transporters, these processes contribute to efficacy of metabolism in both Phase I and II. Drugs used for treatment of cardiovascular disease, such as angiotensin-converting enzyme inhibitors (ACE inhibitors), angiotensin II receptor blockers, diuretics, the aldosterone blocker eplerenone, antiplatelet agents, and oral antithrombotic medications, all show sex differences in efficacy and safety (325, 326). Over-the-counter nonsteroidal anti-inflammatory drugs (NSAIDs) such as ibuprofen and naproxen are more effective in men than women; there is more liver toxicity with acetaminophen use in women, whereas opioids and benzodiazepine work better in women. While some sex differences in metabolic clearance for statins and beta-blockers are known for these frequently prescribed drugs, dosing and adverse event monitoring in routine clinical practice is inadequate. Alosetron, a serotonin receptor 3 antagonist, is approved for treatment of severe irritable bowel syndrome–diarrhea symptoms in women, as it is largely ineffective in men (327). These findings emphasize that women and men take divergent routes (molecular mechanisms and signaling pathways) to reach the same destination (normal function or diseased state), with paths often intersecting. In the era of personalized medicine, there is no one-size-fits-all therapy, and considering sex-specific outcomes in pharmacokinetics and pharmacodynamics of drugs as well as clinical guidelines is warranted to ensure efficacy and safety of medications.

are apparent in all endocrine systems, which are modified by sex chromosomes and sex hormones, with temporal actions across the lifespan.

Blood Pressure Links Cardiovascular and Renal Diseases

Cardiovascular and renal diseases are linked by the relationship of each to arterial pressure (Fig. 5). The cardiovascular system determines arterial pressure, with the heart generating cardiac output and the blood vessels determining total peripheral resistance. The kidneys contribute by regulating extracellular and intravascular fluid volume, and hence blood volume, and venous return. It is established that CVD leads to chronic kidney disease (CKD) and that CKD leads to the development of CVD. For example, following a myocardial infarct, cardiac output declines and arterial pressure falls causing the kidney to vasoconstrict and retain extracellular fluid, with the effect to increase venous return and normalize cardiac output. However, this has the unwanted effect of placing further stress on the failing heart. Conversely, kidney failure causes fluid retention and hypertension (334). Thus, cardiovascular and kidney function are intertwined, as are the endocrine systems that regulate organ function; including the renin-angiotensin-aldosterone system

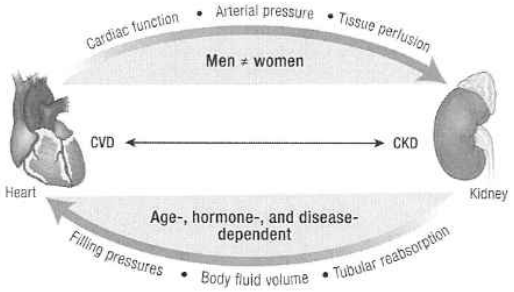


Figure 5. Heart and kidney functions are linked. Sex differences exist in many aspects of heart and kidney function at baseline and in CVD and CKD, as shown. Both organs feed-forward and influence each other's function. Genes, hormones, and age are some known factors that modulate this relationship in a sex-specific manner. Abbreviations: CKD, chronic kidney disease; CVD, cardiovascular disease.

(RAAS), the endothelin system, atrial natriuretic peptides, vasopressin, and glucocorticoid and mineralocorticoid hormones. There is an increasing recognition that there are fundamental sex differences in each of these systems. For example, aldosterone contributes to obesity-induced CVD with a greater impact in females than males (335). However, further research is required to fully elucidate the sex differences present in each endocrine system and how these impact disease development and progression.

Sex Differences in Arterial Pressure and Hypertension

Hypertension is a major risk factor for cardiovascular and renal disease. Over the lifespan there are age- and sex-related differences in arterial pressure. The majority of the data are derived from cross-sectional studies, but a few powerful studies have tracked arterial pressure over decades within a population (332, 336-339). Arterial pressure increases in both men and women with age, although the slope of the relationship is different between men and women. Sex differences in arterial pressure emerge during adolescence and are maintained throughout adulthood until women reach menopause (336, 337, 339). Arterial pressure is ~5 to 10 mmHg greater in men than age-matched women during the reproductive years (340-342). Postmenopause arterial pressure rises steeply in women regardless of race, ethnicity, or country of origin (340-342). One of the most striking characteristics of hypertension is that the prevalence and severity is lower in premenopausal women than in age-matched men. The prevalence of hypertension is ~10% in young premenopausal women, ~50% in postmenopausal women and by the age of 75 years almost ~80% of women are hypertensive (342-344).

Nonhuman mammalian species also display sex differences in arterial pressure. Arterial pressure in adult females is lower in normotensive dogs, sheep, rabbits, rats, and mice as compared with adult males (338, 345). Furthermore, in rodents, rabbits, and sheep, females of reproductive age are protected against the development of hypertension, such that arterial pressure increases significantly less in females than in males, in settings of disease (338). Thus, sex differences are present in the pathophysiology of cardiovascular and renal diseases. Yet, the mechanisms underlying the sexual dimorphism of arterial pressure in men and women as they age are poorly understood. However, extensive evidence indicates that sex hormones likely contribute to the regulation of arterial pressure through their actions on endocrine systems.

Sex Differences in Endocrine Control of Arterial Pressure and Kidney Function

There are subtle differences in most endocrine actions between men and women. It is not the maximal response of each system but rather the slope of the response that is altered. In this manner, a system responds maximally in a hemodynamic crisis (eg, hemorrhage) but in a sex-specific manner to lesser challenges. For example, a greater dose of the vasoconstrictor angiotensin II is required to increase arterial pressure in female than male mice (346). Consistent with this finding, the same dose of angiotensin II caused a greater reduction in renal blood flow in men than women, with the suggestion that this was an angiotensin type 2 receptor (AT_2R) mediated effect (347). In rodents, females of reproductive age have a greater AT_2R to angiotensin type 1 receptor (AT_1R) ratio than males, which contributes to the reduced pressor response to angiotensin II (348). This has been indirectly demonstrated in women, in studies examining forearm vascular resistance responses to AT_2R blockade (349). The AT_2R also mediates a leftward shift in the pressure natriuresis-diuresis relationship, an effect that is greater in female than male mice (350). In women, indirect evidence also indicates a more pronounced role for the AT_2R in the regulation of renal blood flow responses to angiotensin II (347). This is linked to differential expression of components of the RAAS in males and females, which have been demonstrated in most mammalian species, including humans (351). In the context of the above example, estrogen interacts with the glucocorticoid response element on the X-linked *AGTR2* gene, to increase AT_2R expression in females (352). In addition, there are sex differences in human aminopeptidase A, aminopeptidase N, and angiotensin-converting enzyme 2 levels, responsible for generation of the angiotensin peptide fragments, angiotensin III, and angiotensin-(1–7), which have a high affinity for the vasodilatory AT_2R and Mas receptors, respectively (353-356). Lastly, there are marked and important sex differences in the production and function of aldosterone, although this has only recently been started to be examined (335). Thus, in females the RAAS is balanced toward the protective depressor RAAS arm, which at the lower physiological range may prevent arterial pressure increasing to the same extent as in males. However, this delicate balance may be lost in women after menopause and in the situation of metabolic syndrome.

Other vasoconstrictor systems also have sexually dimorphic actions. Endothelin-1 causes vasoconstriction via the endothelin type A receptor (ET_AR), and vasodilation and sodium excretion via the ET_BR. Testosterone increases ET_AR and estrogen increases ET_BR expression, which contributes to the differential control of arterial blood pressure and renal function between the sexes (357). Vasopressin, with important roles in circulatory and water homeostasis, is affected by age and sex. Urinary concentrating ability declines with age, but more steeply in women. Young men produce more concentrated urine than women, in part due higher plasma arginine vasopressin levels and greater vasopressin type 2 receptor expression in the collecting ducts of the kidney in males (358, 359). Renal vasopressin type 2 receptor expression declines with age in association with a reduction in maximal urine concentrating ability (358, 359). Interestingly, aldosterone signaling via mineralocorticoid receptors is associated with increased CVD risk and is

enhanced in obese women (another example of how the RAAS is differentially modulated in females), which has been linked to leptin-induced endothelial dysfunction (360, 361). Moreover, evidence in rodents indicates that sodium reabsorption along the length of the renal tubule is sexually dimorphic, with reabsorption shifted to the later segments in females compared to males. This was associated with greater sodium epithelial channel expression, under the control of aldosterone, in the collecting duct, which could also contribute to the increased cardiovascular and renal risk associated with aldosterone in females (362). Finally, oxytocin, relaxin, and prolactin, which are traditionally known for their roles in pregnancy, have differential cardiovascular and renal actions in nonpregnant female and male rodents (348, 363, 364). Thus, evidence points to sex differences in endocrine control of extracellular fluid homeostasis and vascular function, which likely contribute to age- and sex-related disparities in renal and cardiovascular disease risk. Further studies are warranted to understand this complex issue more fully. In particular, it is important to take into account the subtle effects within the physiological range that counterbalance function of each hormonal system, rather than examine the impact of pharmacological doses which can mask sex differences in responses.

Cardioprotective Mechanisms in Women Sustain a Healthy Pregnancy

The cardioprotective mechanisms that predominate in women during the reproductive years enable the extensive hemodynamic adaptations required to meet the metabolic demands of the developing fetus and a successful pregnancy. During a normotensive pregnancy, blood volume increases and cardiac output increase by ~30% to 50%, but arterial pressure declines due to marked peripheral vasodilatation (365, 366). The associated renal vasodilation accommodates an increase in glomerular filtration rate to process the additional blood volume, but an increase in vasopressin type 2 receptor expression enables increased tubule reabsorption of sodium and water. However, in women with preeclampsia, a pregnancy-induced form of hypertension, these cardiovascular adaptations are perturbed. Accumulating evidence now indicates that women with a history of pregnancy-associated hypertension have a 2- to 5-fold increased risk of CVD in later life (367). Understanding the mechanisms underpinning this dysregulation of vascular function in pregnancy-related hypertension may lead to the identification of new therapeutic targets for the treatment of cardiovascular disease in both sexes. For example, relaxin, which is known best for its role in pregnancy but is also produced in males, plays roles in the regulation of renal function, blood pressure, and tissue fibrosis (363). Thus, it is a mistake to assign hormonal systems a specific role as most have wide-ranging tissue-specific pleiotropic effects.

Sex Hormones and Sex Chromosome Complement in CVD

Sex hormones contribute to sexual dimorphism in endocrine control of the cardiovascular system, with evidence suggesting that there is a "sweet spot" for both testosterone and estradiol, as unusually high or low levels of either promote disease (368-370). This has been the cause of apparent discrepancies in the literature. In particular, this remains a problem in animal studies in which the dose of estrogen used to study the impact of estrogen replacement in aged or gonadectomized models varies widely (~1000-fold), as does the route or length of administration; none of which accurately reflect the cyclic pattern of in vivo production. This lack of rigor into investigation of the effects of sex hormones in preclinical models likely contributes to the controversy that surrounds hormone replacement therapy for the prevention of CVD risk. Despite extensive evidence that hormone replacement therapy is cardioprotective, the negative results of the Women's Health Initiative Trial effectively halted the use of hormone replacement therapy (371). Certainly, high-dose estrogen can increase blood pressure and cardiovascular risk in women (372). However, continued investigation supports the use of hormone replacement therapy in subsets of women, and further work in this area is required (373). In contrast, in men with low testosterone, beneficial cardiovascular effects are seen with testosterone replacement (374). In women with polycystic ovary syndrome, high testosterone levels are associated with elevated blood pressure (374). Dose-ranging studies are required to delineate these effects.

The sex chromosomes may have a direct impact on sex differences in the physiology and pathophysiology of the cardiovascular system and cardiovascular risk, independent of sex hormones. Human sex chromosome aneuploidies, such as Turner and Klinefelter syndromes, suggest that sex chromosome abnormalities can carry an increased risk of CVD. Women with Turner syndrome have around a 3-fold greater mortality and reduced life expectancy relative to the general population (375-377). CVD is a leading cause of increased mortality in Turner syndrome (375-377). Congenital cardiac anomalies, hypertension, coarctation of the aorta, diabetes, ischemic heart disease, and stroke are commonly associated with this condition (378). Similarly, men with Klinefelter syndrome have a high cardiovascular risk profile (379, 380), and an increased risk of

mortality from cardiovascular disease (381, 382). However, observations from studies in individuals with sex chromosome aneuploidies are complicated by confounding factors, including abnormal gonadal sex hormone levels associated with gonadal failure. Thus, it is very difficult to distinguish between hormonal versus genetic mechanisms and cardiovascular risk in these human conditions.

Experimental approaches, such as the FCG mouse model discussed in "Section I," and Box 3 can discriminate between hormonal and sex chromosome effects in cardiovascular disease (115). Beyond genes on the sex chromosomes, there are sex differences in autosomal gene expression, which can be both organ or cell specific (383). In the kidney and the heart, hundreds of rat and human genes are regulated differently between the sexes (384-386). This disparate expression is triggered by sex hormones in ~30% of cases, with the other 70% linked to sex chromosome and microRNA dimorphisms (384, 385). For example, sex differences have been reported in the expression of nitric oxide synthase, tyrosine hydroxylase, and sodium channels in the rodent heart and kidney (332). However, few studies to date have compared gene expression and the effect on the proteome between the human sexes, and further studies are required.

Sex Differences in Pharmacotherapy for Cardiovascular and Renal Disease

Men and women respond to disease differently: kidney diseases progress faster in men than women, kidney transplants from women to men tend to fail more frequently than the reverse, and the effects of diabetes on the kidney differ between the sexes (387-392). Furthermore, symptoms and mechanisms of heart failure differ between the sexes (393). This suggests that sex-specific treatments for CKD and CVD could be required. There is currently little evidence to suggest that men and women respond differently to current treatments for hypertension (394). In large part, this is because clinical trials have lacked statistical power to take this into account. It will be difficult to achieve such an outcome for drugs that have already received FDA approval. However, some treatments are more frequently prescribed, without any basis in evidence (395). There are also marked differences in pharmacokinetics and pharmacodynamics (see Box 4), leading to more frequent adverse drug reactions in women, related to differences in drug clearance and breakdown (396). Therefore, sex should be taken in account for new treatments seeking approval in the future. When women are considered, important and unexpected sex differences are observed in almost every aspect of cardiovascular and renal function in health and disease. Further research is required to fully understand these differences, and in turn to guide the development of sex-specific treatment guidelines for CVD and CKD.

Section V

Challenges for the Future of Sex Differences Research—Areas Requiring Special Attention

Sex differences exist in anatomy, behavior, and physiology across the animal taxa. By extension, because of these innate differences, sex differences exist at molecular and cellular levels in mechanisms that underlie these processes. Despite concerted efforts by the Office of Research on Women's Health and the Organization for the Study of Sex Differences in educating researchers about the distinction between sex versus gender, the indiscriminate use of the word "gender" continues to pervade scientific literature. The sex of established cultured cell lines is another issue; in addition to aneuploidy, chromosomal numbers change as cells are passaged and are dependent upon the tissue of origin (397, 398), but this aspect is beyond the scope of this Statement. Not surprisingly, sex differences are seen in etiology, prevalence, and outcomes in a myriad of human diseases that range from psychological and autoimmune to gastrointestinal, cardiovascular, renal, and reproductive; SARS-CoV-2 causes more severe COVID-19 disease in men than in women despite similar infection rates (399-401). Besides genetic makeup (predisposition), extraneous factors, such as the socioeconomics, demographics, education level, profession, age, and the environment, greatly influence an individual's health; COVID-19 disease outcomes especially highlight the contribution of these extraneous factors in health disparities. Factors such as the endocrine-disruptive chemicals can disproportionately affect one sex over the other; regardless, whether favorable or adverse effects are present in one or both sexes, the effects would impact trans and cisgender persons, and hence these sex-specific effects should not be overlooked or underestimated (402). Some human studies addressing sex differences take these factors into account, whereas others are more selective. Many studies of disease pathways are sensitive to levels of gonadal steroid hormones, which contribute to sex differences. In human studies, unless gender information is explicitly collected or available, the study deals with biological sex, not gender. Use of sex and gender interchangeably deemphasizes the importance of studying gender as an independent variable.

In animals or experimental models of human diseases, effects of estrogens have been investigated more often than effects of progestins and androgens, which should

be corrected. Paradoxically, female sex is often excluded from experimental design on the basis that: (i) the estrus cycle will interfere with data interpretation; (ii) mechanisms that operate in the male sex will operate in the female sex and thus only need to be confirmed in females; (iii) metabolic demands are similar between the sexes; (iv) the X chromosome in males and females is subject to similar regulation; and (v) autosomal genes will be subject to equal variance between the sexes. The same studies often ignore the diurnal cycling nature of testosterone in males; testosterone levels in male rodents can show more day-to-day variability than estrogen and progesterone levels in females. Other steroid hormones, such as glucocorticoids, that show circadian rhythm and whose levels differ between the sexes also influence gene expression and function. In rodents but not primates, sex differences in secretion of growth hormone result in sexually dimorphic hepatic metabolism of drugs and xenobiotics (403). In rodents, endocrine disruption can have transgenerational effects on male and female reproductive systems (404). Since changes in hormone levels and gene expression are dynamic, can be localized, and are spatiotemporally distinct, no one study design or condition can be used as a gold standard. Animal housing and handling conditions can also create sex differences, and thus any experimental design and data interpretation should take these variables into account. If sex-segregated data does not differ for the aspects under study, then data can be pooled from the 2 sexes and reported accordingly.

Studies in animal models have just begun to uncover unequal effects of the sex chromosomes in XX vs XY cells, so we expect further discoveries about such effects in the future. Once genes that cause sex differences are discovered in animals, the findings generate new hypotheses and rationalize human studies to determine whether the same gene also creates sex differences in humans. That question can be studied by the methods of human genetics, relating genetic variation to disease incidence and outcome. Without the animal studies, however, it is difficult to understand detailed molecular mechanisms. It is also important to remember that no single rodent or animal model can capture the complexity of any human disease, but each model provides valuable insights into one or another major aspect of disease. If different etiologies of a given disease share mechanisms, then mimicking the precise conditions that initiate human disease may not be critical.

The study of sex chromosome effects is in its infancy and has focused on proving that sex chromosomes play a role and finding the genes responsible for the effects. So far there has been little effort to understand how these factors interact with steroid hormones to cause sex differences. If both types of factors cause differences in disease incidence, are they affecting the same or different downstream pathways? Do their effects converge, or do they independently affect different mechanisms that each influence a complex disease? Do male-biased factors (hormones, Y-chromosome genes) act synergistically to induce a male-specific state, or do they counteract each other to reduce the difference between males and females (123, 405)? Are the diverse sex-biasing factors changing in their effects across the lifespan, leading to changes in the type or amount of sex difference at different ages?

When studying sex differences in animal models of human diseases, it is important to first understand and elucidate differences at baseline in gonadally intact animals. As pointed out earlier, steroidogenic enzymes are also present in nongonadal tissues, especially the brain, thus it is not entirely possible to eliminate effects of sex steroids from all tissues. Moreover, tamoxifen-inducible *Cre* recombinase used to routinely perform lineage tracing and gene inactivation studies in mice has its own problems (406, 407) that are largely ignored and can further confound sex-specific data analysis; tamoxifen antagonizes actions of estrogen receptor-β and inhibits expression of over 70 genes (408), but the contribution of these tamoxifen-regulated genes on study results and outcomes is never accounted for and requires careful consideration. Before mechanisms behind sex differences in physiology and disease can be elucidated, a fundamental understanding of sex differences that exist at baseline, is needed.

Acknowledgments

The authors thank Stephen M. Rosenthal and Robert M. Carey for critically reading the manuscript.

Additional Information

Correspondence: Aditi Bhargava, PhD, Professor, Department of ObGyn and Center for Reproductive Sciences, 513 Parnassus Avenue, HSE1635, Box 0556, UCSF, San Francisco, CA 94143, USA. Email: Aditi.bhargava@ucsf.edu

Disclosures: The authors have nothing to disclose.

References

1. Baclawski K. The Observer Effect. *2018 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA). Boston, MA,* 2018:83-89. doi:10.1109/COGSIMA.2018.8423983.
2. Garnjobst L, Wilson JF. Heterocaryosis and protoplasmic incompatibility in neurospora crassa. *Proc Natl Acad Sci U S A.* 1956;**42**(9):613-618.
3. Hadjivasiliou Z, Pomiankowski A. Evolution of asymmetric gamete signaling and suppressed recombination at the mating type locus. *Elife.* 2019;**8**:e48239.
4. Martin CL, Ruble DN. Patterns of gender development. *Annu Rev Psychol.* 2010;**61**:353-381.
5. van Anders SM, Goldey KL, Kuo PX. The steroid/peptide theory of social bonds: integrating testosterone and peptide responses for classifying social behavioral contexts. *Psychoneuroendocrinology.* 2011;**36**(9):1265-1275.
6. Ellegren H. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet.* 2011;**12**(3):157-166.
7. Matsubara K, Tarui H, Toriba M, et al. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *Proc Natl Acad Sci U S A.* 2006;**103**(48):18190-18195.
8. Munday PL, Buston PM, Warner RR. Diversity and flexibility of sex-change strategies in animals. *Trends Ecol Evol.* 2006;**21**(2):89-95.
9. Todd EV, Ortega-Recalde O, Liu H, et al. Stress, novel sex genes, and epigenetic reprogramming orchestrate socially controlled sex change. *Sci Adv.* 2019;**5**:eaaw7006.
10. Matson CK, Murphy MW, Sarver AL, Griswold MD, Bardwell VJ, Zarkower D. DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature.* 2011;**476**(7358):101-104.
11. Uhlenhaut NH, Jakob S, Anlag K, et al. Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell.* 2009;**139**(6):1130-1142.
12. Berta P, Hawkins JR, Sinclair AH, et al. Genetic evidence equating SRY and the testis-determining factor. *Nature.* 1990;**348**(6300):448-450.
13. Sinclair AH, Berta P, Palmer MS, et al. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature.* 1990;**346**(6281):240-244.
14. Fechner PY, Marcantonio SM, Jaswaney V, et al. The role of the sex-determining region Y gene in the etiology of 46,XX maleness. *J Clin Endocrinol Metab.* 1993;**76**(3):690-695.
15. Vorona E, Zitzmann M, Gromoll J, Schüring AN, Nieschlag E. Clinical, endocrinological, and epigenetic features of the 46,XX male syndrome, compared with 47,XXY Klinefelter patients. *J Clin Endocrinol Metab.* 2007;**92**(9):3458-3465.
16. Ludbrook LM, Harley VR. Sex determination: a 'window' of DAX1 activity. *Trends Endocrinol Metab.* 2004;**15**(3):116-121.
17. Arnold AP. A general theory of sexual differentiation. *J Neurosci Res.* 2017;**95**:291-300.
18. Spiller C, Koopman P, Bowles J. Sex determination in the mammalian germline. *Annu Rev Genet.* 2017;**51**:265-285.
19. Edson MA, Nagaraja AK, Matzuk MM. The mammalian ovary from genesis to revelation. *Endocr Rev.* 2009;30(6):624-712.
20. Eid W, Biason-Lauber A. Why boys will be boys and girls will be girls: human sex development and its defects. *Birth Defects Res C Embryo Today.* 2016;**108**(4):365-379.
21. Parma P, Radi O, Vidal V, et al. R-spondin1 is essential in sex determination, skin differentiation and malignancy. *Nat Genet.* 2006;**38**(11):1304-1309.
22. Zhao F, Franco HL, Rodriguez KF, et al. Elimination of the male reproductive tract in the female embryo is promoted by COUP-TFII in mice. *Science.* 2017;**357**(6352):717-720.
23. Burgoyne PS, Mitchell MJ. The role of mouse Y chromosome genes in spermatogenesis. In: Lau YFC, Chan WY, eds. *Y Chromosome and Male Germ Cell Biology.* Hackensack NJ: World Scientific Publishers; 2007:27-45.
24. Hughes JF, Page DC. The biology and evolution of mammalian Y chromosomes. *Annu Rev Genet.* 2015;**49**:507-527.
25. Lyon MF. Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet.* 1962;**14**:135-148.
26. Russell LB. Mammalian X-chromosome action: inactivation limited in spread and region of origin. *Science.* 1963;**140**(3570):976-978.
27. Brown CJ, Ballabio A, Rupert JL, et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature.* 1991;**349**(6304):38-44.
28. Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell.* 2013;**152**(6):1308-1323.
29. Disteche CM. Dosage compensation of the sex chromosomes and autosomes. *Semin Cell Dev Biol.* 2016;**56**:9-18.
30. Tukiainen T, Villani AC, Yen A, et al. Landscape of X chromosome inactivation across human tissues. *Nature.* 2017;**550**(7675):244-248.
31. Carrel L, Cottle AA, Goglin KC, Willard HF. A first-generation X-inactivation profile of the human X chromosome. *Proc Natl Acad Sci U S A.* 1999;**96**(25):14440-14444.
32. Nugent BM, O'Donnell CM, Epperson CN, Bale TL. Placental H3K27me3 establishes female resilience to prenatal insults. *Nat Commun.* 2018;**9**(1):2555.
33. Kaneko S, Li X. X chromosome protects against bladder cancer in females via a KDM6A-dependent epigenetic mechanism. *Sci Adv.* 2018;**4**:eaar5598.
34. Itoh Y, Golden LC, Itoh N, et al. The X-linked histone demethylase Kdm6a in CD4+ T lymphocytes modulates autoimmunity. *J Clin Invest.* 2019;**129**(9):3852-3863.
35. Davis EJ, Broestl L, Abdulai-Saiku S, et al. A second X chromosome contributes to resilience in a mouse model of Alzheimer's disease. *Sci Transl Med.* 2020;**12**(558):eaaz5677.
36. Link JC, Wiese CB, Chen X, et al. X chromosome dosage of histone demethylase KDM5C determines sex differences in adiposity. *J Clin Invest.* 2020;**130**(11):5688-5702.
37. Golden LC, Itoh Y, Itoh N, et al. Parent-of-origin differences in DNA methylation of X chromosome genes in T lymphocytes. *Proc Natl Acad Sci U S A.* 2019;**116**:26779-26787.
38. Migeon BR. Why females are mosaics, X-chromosome inactivation, and sex differences in disease. *Gend Med.* 2007;**4**(2):97-105.
39. MacLaughlin DT, Donahoe PK. Sex determination and differentiation. *N Engl J Med.* 2004;**350**(4):367-378.
40. Voutilainen R, Miller WL. Developmental expression of genes for the stereoidogenic enzymes P450scc (20,22-desmolase), P450c17 (17α-hydroxylase/17,20-lyase), and P450c21 (21-hydroxylase) in the human fetus. *J Clin Endocrinol Metab.* 1986;**63**(5):1145-1150.

41. Jost A. On the effects of early castration of the male rabbit embryo. *C R Seances Soc Biol Fil.* 1947;**141**(3-4):126-129.
42. Jost A, Vigier B, Prépin J, Perchellet JP. Studies on sex differentiation in mammals. *Recent Prog Horm Res.* 1973;**29**:1-41.
43. Ingraham HA, Hirokawa Y, Roberts LM, et al. Autocrine and paracrine Müllerian inhibiting substance hormone signaling in reproduction. *Recent Prog Horm Res.* 2000;**55**:53-67; discussion 67.
44. Lane AH, Donahoe PK. New insights into mullerian inhibiting substance and its mechanism of action. *J Endocrinol.* 1998;**158**(1):1-6.
45. Swain A. Ductal sex determination. *Science.* 2017;**357**(6352):648.
46. Flück CE, Meyer-Böni M, Pandey AV, et al. Why boys will be boys: two pathways of fetal testicular androgen biosynthesis are needed for male sexual differentiation. *Am J Hum Genet.* 2011;**89**(2):201-218.
47. Arlt W, Martens JW, Song M, Wang JT, Auchus RJ, Miller WL. Molecular evolution of adrenarche: structural and functional analysis of p450c17 from four primate species. *Endocrinology.* 2002;**143**(12):4665-4672.
48. Cutler GB Jr, Glenn M, Bush M, Hodgen GD, Graham CE, Loriaux DL. Adrenarche: a survey of rodents, domestic animals, and primates. *Endocrinology.* 1978;**103**(6):2112-2118.
49. Becker M, Hesse V. Minipuberty: why does it happen? *Horm Res Paediatr.* 2020;1-10.
50. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev.* 2011;**32**(1):81-151.
51. Miller WL, Bose HS. Early steps in steroidogenesis: intracellular cholesterol trafficking. *J Lipid Res.* 2011;**52**(12):2111-2135.
52. Mellon SH, Griffin LD. Neurosteroids: biochemistry and clinical significance. *Trends Endocrinol Metab.* 2002;**13**(1):35-43.
53. Missaghian E, Kempná P, Dick B, et al. Role of DNA methylation in the tissue-specific expression of the CYP17A1 gene for steroidogenesis in rodents. *J Endocrinol.* 2009;**202**(1):99-109.
54. Rege J, Turcu AF, Kasa-Vubu JZ, et al. 11-Ketotestosterone is the dominant circulating bioactive androgen during normal and premature adrenarche. *J Clin Endocrinol Metab.* 2018;**103**(12):4589-4598.
55. Nyby JG. Reflexive testosterone release: a model system for studying the nongenomic effects of testosterone upon male behavior. *Front Neuroendocrinol.* 2008;**29**(2):199-210.
56. Brouillette J, Rivard K, Lizotte E, Fiset C. Sex and strain differences in adult mouse cardiac repolarization: importance of androgens. *Cardiovasc Res.* 2005;**65**(1):148-157.
57. Nilsson ME, Vandenput L, Tivesten Å, et al. Measurement of a comprehensive sex steroid profile in rodent serum by high-sensitive gas chromatography-tandem mass spectrometry. *Endocrinology.* 2015;**156**(7):2492-2502.
58. Laurent MR, Hammond GL, Blokland M, et al. Sex hormone-binding globulin regulation of androgen bioactivity in vivo: validation of the free hormone hypothesis. *Sci Rep.* 2016;**6**:35539.
59. van Weerden WM, Bierings HG, van Steenbrugge GJ, de Jong FH, Schröder FH. Adrenal glands of mouse and rat do not synthesize androgens. *Life Sci.* 1992;**50**(12):857-861.
60. Morris JA, Jordan CL, Breedlove SM. Sexual differentiation of the vertebrate nervous system. *Nat Neurosci.* 2004;**7**(10):1034-1039.
61. McCarthy MM, Herold K, Stockman SL. Fast, furious and enduring: sensitive versus critical periods in sexual differentiation of the brain. *Physiol Behav.* 2018;**187**:13-19.
62. Bakker J, Brock O. Early oestrogens in shaping reproductive networks: evidence for a potential organisational role of oestradiol in female brain development. *J Neuroendocrinol.* 2010;**22**(7):728-735.
63. Roselli CE, Klosterman SA. Sexual differentiation of aromatase activity in the rat brain: effects of perinatal steroid exposure. *Endocrinology.* 1998;**139**(7):3193-3201.
64. McCarthy MM. Estradiol and the developing brain. *Physiol Rev.* 2008;**88**(1):91-124.
65. Bakker J, De Mees C, Douhard Q, et al. Alpha-fetoprotein protects the developing female mouse brain from masculinization and defeminization by estrogens. *Nat Neurosci.* 2006;**9**(2):220-226.
66. Hammond GL. Access of reproductive steroids to target tissues. *Obstet Gynecol Clin North Am.* 2002;**29**(3):411-423.
67. Hong H, Branham WS, Ng HW, et al. Human sex hormone-binding globulin binding affinities of 125 structurally diverse chemicals and comparison with their binding to androgen receptor, estrogen receptor, and α-fetoprotein. *Toxicol Sci.* 2015;**143**(2):333-348.
68. Aussel C, Masseyeff R. Comparative binding properties of rat and human alpha-fetoproteins for arachidonic acid and estradiol. *Res Commun Chem Pathol Pharmacol.* 1983;**42**(2):261-269.
69. Mizejewski GJ. Biological roles of alpha-fetoprotein during pregnancy and perinatal development. *Exp Biol Med (Maywood).* 2004;**229**(6):439-463.
70. Breedlove SM. Sexual differentiation of the human nervous system. *Annu Rev Psychol.* 1994;**45**:389-418.
71. Place NJ, Holekamp KE, Sisk CL, et al. Effects of prenatal treatment with antiandrogens on luteinizing hormone secretion and sex steroid concentrations in adult spotted hyenas, Crocuta crocuta. *Biol Reprod.* 2002;**67**(5):1405-1413.
72. Amateau SK, McCarthy MM. Induction of PGE2 by estradiol mediates developmental masculinization of sex behavior. *Nat Neurosci.* 2004;**7**(6):643-650.
73. Petersen SL, Krishnan S, Aggison LK, Intlekofer KA, Moura PJ. Sexual differentiation of the gonadotropin surge release mechanism: a new role for the canonical NfκB signaling pathway. *Front Neuroendocrinol.* 2012;**33**(1):36-44.
74. Witchel SF. Disorders of sex development. *Best Pract Res Clin Obstet Gynaecol.* 2018;**48**:90-102.
75. Lee PA, Houk CP, Ahmed SF, Hughes IA; International Consensus Conference on Intersex organized by the Lawson Wilkins Pediatric Endocrine Society and the European Society for Paediatric Endocrinology. Consensus statement on management of intersex disorders. International Consensus Conference on Intersex. *Pediatrics.* 2006;**118**(2):e488-e500.
76. Baetens D, Verdin H, De Baere E, Cools M. Update on the genetics of differences of sex development (DSD). *Best Pract Res Clin Endocrinol Metab.* 2019;**33**(3):101271.
77. Miller WL. Disorders in the initial steps of steroid hormone synthesis. *J Steroid Biochem Mol Biol.* 2017;**165**(Pt A):18-37.
78. 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.

79. Martínez-Frías ML. Developmental field defects and associations: epidemiological evidence of their relationship. *Am J Med Genet*. 1994;**49**(1):45-51.
80. Fontana L, Gentilin B, Fedele L, Gervasini C, Miozzo M. Genetics of Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome. *Clin Genet*. 2017;**91**(2):233-246.
81. Hiort O, Birnbaum W, Marshall L, et al. Management of disorders of sex development. *Nat Rev Endocrinol*. 2014;**10**(9):520-529.
82. Achermann JC, Domenice S, Bachega TA, Nishi MY, Mendonca BB. Disorders of sex development: effect of molecular diagnostics. *Nat Rev Endocrinol*. 2015;**11**(8):478-488.
83. Sandberg DE, Gardner M, Callens N, Mazur T; DSD-TRN Psychosocial Workgroup, the DSD-TRN Advocacy Advisory Network, and Accord Alliance. Interdisciplinary care in disorders/differences of sex development (DSD): the psychosocial component of the DSD-Translational research network. *Am J Med Genet C Semin Med Genet*. 2017;**175**(2):279-292.
84. Hembree WC, Cohen-Kettenis PT, Gooren L, et al. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2017;**102**(11):3869-3903.
85. Winter S, Diamond M, Green J, et al. Transgender people: health at the margins of society. *Lancet*. 2016;**388**(10042):390-400.
86. Johns MM, Lowry R, Andrzejewski J, et al. Transgender identity and experiences of violence victimization, substance use, suicide risk, and sexual risk behaviors among high school students - 19 states and large urban school districts, 2017. *MMWR Morb Mortal Wkly Rep*. 2019;**68**:67-71.
87. Rider GN, McMorris BJ, Gower AL, Coleman E, Eisenberg ME. Health and care utilization of transgender and gender nonconforming youth: a population-based study. *Pediatrics*. 2018;**141**(3):e20171683.
88. Herman JL, Flores AR, Brown TNT, Wilson BDM, Conron KJ. *Age of Individuals Who Identify as Transgender in the United States*. Los Angeles: The Williams Institute, UCLA School of Law; 2017.
89. Rosenthal SM. Approach to the patient: transgender youth: endocrine considerations. *J Clin Endocrinol Metab*. 2014;**99**(12):4379-4389.
90. Saraswat A, Weinand JD, Safer JD. Evidence supporting the biologic nature of gender identity. *Endocr Pract*. 2015;**21**(2):199-204.
91. Steensma TD, Kreukels BP, de Vries AL, Cohen-Kettenis PT. Gender identity development in adolescence. *Horm Behav*. 2013;**64**(2):288-297.
92. Allen LS, Hines M, Shryne JE, Gorski RA. Two sexually dimorphic cell groups in the human brain. *J Neurosci*. 1989;**9**(2):497-506.
93. LeVay S. A difference in hypothalamic structure between heterosexual and homosexual men. *Science*. 1991;**253**:1034-1037.
94. Byne W, Tobet S, Mattiace LA, et al. The interstitial nuclei of the human anterior hypothalamus: an investigation of variation with sex, sexual orientation, and HIV status. *Horm Behav*. 2001;**40**(2):86-92.
95. Kreukels BP, Guillamon A. Neuroimaging studies in people with gender incongruence. *Int Rev Psychiatry*. 2016;**28**(1):120-128.
96. Feusner JD, Lidström A, Moody TD, Dhejne C, Bookheimer SY, Savic I. Intrinsic network connectivity and own body perception in gender dysphoria. *Brain Imaging Behav*. 2017;**11**(4):964-976.
97. Hahn A, Kranz GS, Sladky R, et al. Testosterone affects language areas of the adult human brain. *Hum Brain Mapp*. 2016;**37**(5):1738-1748.
98. Luders E, Sánchez FJ, Tosun D, et al. Increased cortical thickness in male-to-female transsexualism. *J Behav Brain Sci*. 2012;**2**(3):357-362.
99. Burke SM, Manzouri AH, Dhejne C, et al. Testosterone effects on the brain in transgender men. *Cereb Cortex*. 2018;**28**(5):1582-1596.
100. Coolidge FL, Thede LL, Young SE. The heritability of gender identity disorder in a child and adolescent twin sample. *Behav Genet*. 2002;**32**(4):251-257.
101. Heylens G, De Cuypere G, Zucker KJ, et al. Gender identity disorder in twins: a review of the case report literature. *J Sex Med*. 2012;**9**(3):751-757.
102. Berenbaum SA, Meyer-Bahlburg HF. Gender development and sexuality in disorders of sex development. *Horm Metab Res*. 2015;**47**(5):361-366.
103. Dessens AB, Slijper FM, Drop SL. Gender dysphoria and gender change in chromosomal females with congenital adrenal hyperplasia. *Arch Sex Behav*. 2005;**34**(4):389-397.
104. Frisén L, Nordenström A, Falhammar H, et al. Gender role behavior, sexuality, and psychosocial adaptation in women with congenital adrenal hyperplasia due to CYP21A2 deficiency. *J Clin Endocrinol Metab*. 2009;**94**(9):3432-3439.
105. Meyer-Bahlburg HF, Dolezal C, Baker SW, Carlson AD, Obeid JS, New MI. Prenatal androgenization affects gender-related behavior but not gender identity in 5-12-year-old girls with congenital adrenal hyperplasia. *Arch Sex Behav*. 2004;**33**(2):97-104.
106. Cohen-Kettenis PT. Gender change in 46,XY persons with 5α-reductase-2 deficiency and 17β-hydroxysteroid dehydrogenase-3 deficiency. *Arch Sex Behav*. 2005;**34**(4):399-410.
107. Fisher AD, Ristori J, Morelli G, Maggi M. The molecular mechanisms of sexual orientation and gender identity. *Mol Cell Endocrinol*. 2018;**467**:3-13.
108. Lentini E, Kasahara M, Arver S, Savic I. Sex differences in the human brain and the impact of sex chromosomes and sex hormones. *Cereb Cortex*. 2013;**23**(10):2322-2336.
109. Raznahan A, Lee NR, Greenstein D, et al. Globally divergent but locally convergent X- and Y-chromosome influences on cortical development. *Cereb Cortex*. 2016;**26**(1):70-79.
110. Raznahan A, Parikshak NN, Chandran V, et al. Sex-chromosome dosage effects on gene expression in humans. *Proc Natl Acad Sci U S A*. 2018;**115**(28):7398-7403.
111. Savic I, Frisen L, Manzouri A, Nordenstrom A, Lindén Hirschberg A. Role of testosterone and Y chromosome genes for the masculinization of the human brain. *Hum Brain Mapp*. 2017;**38**(4):1801-1814.
112. Arnold AP, Chen X. What does the "four core genotypes" mouse model tell us about sex differences in the brain and other tissues? *Front Neuroendocrinol*. 2009;**30**(1):1-9.
113. Arnold AP. Conceptual frameworks and mouse models for studying sex differences in physiology and disease: why compensation changes the game. *Exp Neurol*. 2014;**259**:2-9.
114. Burgoyne PS, Arnold AP. A primer on the use of mouse models for identifying direct sex chromosome effects that cause sex differences in non-gonadal tissues. *Biol Sex Differ*. 2016;**7**:68.
115. Arnold AP, Cassis LA, Eghbali M, Reue K, Sandberg K. Sex hormones and sex chromosomes cause sex differences in the

development of cardiovascular diseases. *Arterioscler Thromb Vasc Biol.* 2017;**37**(5):746-756.
116. Arnold AP, Reue K, Eghbali M, et al. The importance of having two X chromosomes. *Philos Trans R Soc Lond B Biol Sci.* 2016;**371**(1688):20150113.
117. Umar S, Cunningham CM, Itoh Y, et al. The Y chromosome plays a protective role in experimental hypoxic pulmonary hypertension. *Am J Respir Crit Care Med.* 2018;**197**(7):952-955.
118. Shpargel KB, Sengoku T, Yokoyama S, Magnuson T. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *Plos Genet.* 2012;**8**(9):e1002964.
119. Li J, Chen X, McClusky R, et al. The number of X chromosomes influences protection from cardiac ischaemia/reperfusion injury in mice: one X is better than two. *Cardiovasc Res.* 2014;**102**(3):375-384.
120. Ji H, Zheng W, Wu X, et al. Sex chromosome effects unmasked in angiotensin II-induced hypertension. *Hypertension.* 2010;**55**(5):1275-1282.
121. Arnold AP. The end of gonad-centric sex determination in mammals. *Trends Genet.* 2012;**28**(2):55-61.
122. Lowe R, Gemma C, Rakyan VK, Holland ML. Sexually dimorphic gene expression emerges with embryonic genome activation and is dynamic throughout development. *BMC Genomics.* 2015;**16**:295.
123. Arnold AP. Rethinking sex determination of non-gonadal tissues. *Curr Top Dev Biol.* 2019;**134**:289-315.
124. Shansky RM. Are hormones a "female problem" for animal research? *Science.* 2019;**364**(6443):825-826.
125. Exploring the biological contributions to human health: does sex matter? *J Womens Health Gend Based Med.* 2001;**10**:433-439.
126. T'Sjoen G, Arcelus J, Gooren L, Klink DT, Tangpricha V. Endocrinology of transgender medicine. *Endocr Rev.* 2019;**40**(1):97-117.
127. Hasdemir B, Mhaske P, Paruthiyil S, et al. Sex- and corticotropin-releasing factor receptor 2- dependent actions of urocortin 1 during inflammation. *Am J Physiol Regul Integr Comp Physiol.* 2016;**310**(11):R1244-R1257.
128. Kubat E, Mahajan S, Liao M, et al. Corticotropin-releasing factor receptor 2 mediates sex-specific cellular stress responses. *Mol Med.* 2013;**19**:212-222.
129. Paruthiyil S, Hagiwara SI, Kundassery K, Bhargava A. Sexually dimorphic metabolic responses mediated by CRF2 receptor during nutritional stress in mice. *Biol Sex Differ.* 2018;**9**(1):49.
130. Sorge RE, Martin LJ, Isbester KA, et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat Methods.* 2014;**11**(6):629-632.
131. Chen P, Hong W. Neural circuit mechanisms of social behavior. *Neuron.* 2018;**98**(1):16-30.
132. Brown GR, Laland KN, Mulder MB. Bateman's principles and human sex roles. *Trends Ecol Evol.* 2009;**24**(6):297-304.
133. Archer J. Does sexual selection explain human sex differences in aggression? *Behav Brain Sci.* 2009;**32**:249-266; discussion 266-311.
134. Rutter M, Caspi A, Moffitt TE. Using sex differences in psychopathology to study causal mechanisms: unifying issues and research strategies. *J Child Psychol Psychiatry.* 2003;**44**(8):1092-1115.
135. Crossley NA, Fox PT, Bullmore ET. Meta-connectomics: human brain network and connectivity meta-analyses. *Psychol Med.* 2016;**46**(5):897-907.
136. Giedd JN, Raznahan A, Mills KL, Lenroot RK. Review: magnetic resonance imaging of male/female differences in human adolescent brain anatomy. *Biol Sex Differ.* 2012;**3**(1):19.
137. Qiu LR, Fernandes DJ, Szulc-Lerch KU, et al. Mouse MRI shows brain areas relatively larger in males emerge before those larger in females. *Nat Commun.* 2018;**9**(1):2615.
138. Swanson LW, Bota M. Foundational model of structural connectivity in the nervous system with a schema for wiring diagrams, connectome, and basic plan architecture. *Proc Natl Acad Sci U S A.* 2010;**107**(48):20610-20617.
139. Chen K, Azeez A, Chen DY, Biswal BB. Resting-State Functional Connectivity: Signal Origins and Analytic Methods. *Neuroimaging Clin N Am.* 2020;**30**(1):15-23.
140. Smith SM, Vidaurre D, Beckmann CF, et al. Functional connectomics from resting-state fMRI. *Trends Cogn Sci.* 2013;**17**(12):666-682.
141. Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J Magn Reson B.* 1996;**111**(3):209-219.
142. Hagmann P, Cammoun L, Gigandet X, et al. Mapping the structural core of human cerebral cortex. *Plos Biol.* 2008;**6**(7):e159.
143. Mori S, Crain BJ, Chacko VP, van Zijl PC. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann Neurol.* 1999;**45**(2):265-269.
144. Savic I, Arver S. Sex differences in cortical thickness and their possible genetic and sex hormonal underpinnings. *Cereb Cortex.* 2014;**24**(12):3246-3257.
145. Gerhard S, Daducci A, Lemkaddem A, Meuli R, Thiran JP, Hagmann P. The connectome viewer toolkit: an open source framework to manage, analyze, and visualize connectomes. *Front Neuroinform.* 2011;**5**:3.
146. Glasser MF, Smith SM, Marcus DS, et al. The Human Connectome Project's neuroimaging approach. *Nat Neurosci.* 2016;**19**(9):1175-1187.
147. Shen X, Finn ES, Scheinost D, et al. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat Protoc.* 2017;**12**(3):506-518.
148. Lotze M, Domin M, Gerlach FH, et al. Novel findings from 2838 adult brains on sex differences in gray matter brain volume. *Sci Rep.* 2019;**9**(1):1671.
149. Ritchie SJ, Cox SR, Shen X, et al. Sex differences in the adult human brain: evidence from 5216 UK Biobank participants. *Cereb Cortex.* 2018;**28**(8):2959-2975.
150. Knickmeyer RC, Xia K, Lu Z, et al. Impact of demographic and obstetric factors on infant brain volumes: a population neuroscience study. *Cereb Cortex.* 2017;**27**(12):5616-5625.
151. Lenroot RK, Gogtay N, Greenstein DK, et al. Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *Neuroimage.* 2007;**36**(4):1065-1073.
152. Giedd JN, Raznahan A, Alexander-Bloch A, Schmitt E, Gogtay N, Rapoport JL. Child psychiatry branch of the National Institute of Mental Health longitudinal structural

magnetic resonance imaging study of human brain development. *Neuropsychopharmacology*. 2015;**40**(1):43-49.
153. Kuczmarski RJ, Ogden CL, Guo SS, et al. 2000 CDC growth charts for the United States: methods and development. *Vital Health Stat 11*. 2002;1-190.
154. Mutlu AK, Schneider M, Debbané M, Badoud D, Eliez S, Schaer M. Sex differences in thickness, and folding developments throughout the cortex. *Neuroimage*. 2013;**82**:200-207.
155. Liu S, Seidlitz J, Blumenthal JD, Clasen LS, Raznahan A. Integrative structural, functional, and transcriptomic analyses of sex-biased brain organization in humans. *Proc Natl Acad Sci U S A*. 2020;**117**(31):18788-18798.
156. Ruigrok AN, Salimi-Khorshidi G, Lai MC, et al. A meta-analysis of sex differences in human brain structure. *Neurosci Biobehav Rev*. 2014;**39**:34-50.
157. Raznahan A, Lue Y, Probst F, et al. Triangulating the sexually dimorphic brain through high-resolution neuroimaging of murine sex chromosome aneuploidies. *Brain Struct Funct*. 2015;**220**(6):3581-3593.
158. Fish AM, Nadig A, Seidlitz J, et al. Sex-biased trajectories of amygdalo-hippocampal morphology change over human development. *Neuroimage*. 2020;**204**:116122.
159. Herting MM, Maxwell EC, Irvine C, Nagel BJ. The impact of sex, puberty, and hormones on white matter microstructure in adolescents. *Cereb Cortex*. 2012;**22**(9):1979-1992.
160. Sporns O, Tononi G, Kötter R. The human connectome: a structural description of the human brain. *Plos Comput Biol*. 2005;**1**(4):e42.
161. Tunc B, Solmaz B, Parker D, et al. Establishing a link between sex-related differences in the structural connectome and behaviour. *Philos Trans R Soc Lond B Biol Sci*. 2016;**371**(1688):20150111.
162. Ingalhalikar M, Smith A, Parker D, et al. Sex differences in the structural connectome of the human brain. *Proc Natl Acad Sci U S A*. 2014;**111**(2):823-828.
163. Gur RC, Richard J, Calkins ME, et al. Age group and sex differences in performance on a computerized neurocognitive battery in children age 8–21. *Neuropsychology*. 2012;**26**(2):251-265.
164. Zhang C, Cahill ND, Arbabshirani MR, White T, Baum SA, Michael AM. Sex and age effects of functional connectivity in early adulthood. *Brain Connect*. 2016;**6**.
165. Girvan M, Newman ME. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*. 2002;**99**(12):7821-7826.
166. Tononi G, Sporns O, Edelman GM. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc Natl Acad Sci U S A*. 1994;**91**(11):5033-5037.
167. Schwarz AJ, Gozzi A, Bifone A. Community structure and modularity in networks of correlated brain activity. *Magn Reson Imaging*. 2008;**26**(7):914-920.
168. Moreno-Briseño P, Díaz R, Campos-Romo A, Fernandez-Ruiz J. Sex-related differences in motor learning and performance. *Behav Brain Funct*. 2010;**6**:74.
169. Hedges LV, Nowell A. Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*. 1995;**269**(5220):41-45.
170. Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol Bull*. 1995;**117**(2):250-270.
171. Maguire EA, Gadian DG, Johnsrude IS, et al. Navigation-related structural change in the hippocampi of taxi drivers. *Proc Natl Acad Sci U S A*. 2000;**97**(8):4398-4403.
172. Maguire EA, Woollett K, Spiers HJ. London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. *Hippocampus*. 2006;**16**(12):1091-1101.
173. Draganski B, Gaser C, Busch V, Schuierer G, Bogdahn U, May A. Neuroplasticity: changes in grey matter induced by training. *Nature*. 2004;**427**(6972):311-312.
174. Gaser C, Schlaug G. Gray matter differences between musicians and nonmusicians. *Ann N Y Acad Sci*. 2003;**999**:514-517.
175. Mechelli A, Crinion JT, Noppeney U, et al. Neurolinguistics: structural plasticity in the bilingual brain. *Nature*. 2004;**431**(7010):757.
176. Schlegel AA, Rudelson JJ, Tse PU. White matter structure changes as adults learn a second language. *J Cogn Neurosci*. 2012;**24**(8):1664-1670.
177. Clayton JA. Applying the new SABV (sex as a biological variable) policy to research and clinical care. *Physiol Behav*. 2018;**187**:2-5.
178. Hallam J, Boswell RG, DeVito EE, Kober H. Gender-related differences in food craving and obesity. *Yale J Biol Med*. 2016;**89**(2):161-173.
179. Striegel-Moore RH, Rosselli F, Perrin N, et al. Gender difference in the prevalence of eating disorder symptoms. *Int J Eat Disord*. 2009;**42**(5):471-474.
180. Pursey KM, Stanwell P, Callister RJ, Brain K, Collins CE, Burrows TL. Neural responses to visual food cues according to weight status: a systematic review of functional magnetic resonance imaging studies. *Front Nutr*. 2014;**1**:7.
181. Cepeda-Benito A, Fernandez MC, Moreno S. Relationship of gender and eating disorder symptoms to reported cravings for food: construct validation of state and trait craving questionnaires in Spanish. *Appetite*. 2003;**40**(1):47-54.
182. Imperatori C, Innamorati M, Tamburello S, et al. Gender differences in food craving among overweight and obese patients attending low energy diet therapy: a matched case-control study. *Eat Weight Disord*. 2013;**18**(3):297-303.
183. Zellner DA, Garriga-Trillo A, Rohm E, Centeno S, Parker S. Food liking and craving: a cross-cultural approach. *Appetite*. 1999;**33**(1):61-70.
184. Lafay L, Thomas F, Mennen L, et al.; Fleurbaix Laventie Ville Santé Study Group. Gender differences in the relation between food cravings and mood in an adult community: results from the fleurbaix laventie ville santé study. *Int J Eat Disord*. 2001;**29**(2):195-204.
185. Pelchat ML. Food cravings in young and elderly adults. *Appetite*. 1997;**28**(2):103-113.
186. Wang GJ, Volkow ND, Telang F, et al. Evidence of gender differences in the ability to inhibit brain activation elicited by food stimulation. *Proc Natl Acad Sci U S A*. 2009;**106**(4):1249-1254.
187. Croll J, Neumark-Sztainer D, Story M, Ireland M. Prevalence and risk and protective factors related to disordered eating behaviors among adolescents: relationship to gender and ethnicity. *J Adolesc Health*. 2002;**31**(2):166-175.
188. Kenny PJ. Reward mechanisms in obesity: new insights and future directions. *Neuron*. 2011;**69**(4):664-679.
189. Saper CB, Chou TC, Elmquist JK. The need to feed: homeostatic and hedonic control of eating. *Neuron*. 2002;**36**(2):199-211.

190. Zheng H, Berthoud HR. Eating for pleasure or calories. *Curr Opin Pharmacol*. 2007;7(6):607-612.
191. Ziauddeen H, Alonso-Alonso M, Hill JO, Kelley M, Khan NA. Obesity and the neurocognitive basis of food reward and the control of intake. *Adv Nutr*. 2015;6(4):474-486.
192. Gupta AM, EA, Sanmiguel CP, et al. Patterns of brain structural connectivity differentiate normal weight from overweight subjects. *Neuroimage Clin*. 2015;7:506-517.
193. Yokum SN, J, Stice E. Attentional bias to food images associated with elevated weight and future weight gain: an fMRI study. *Obesity (Silver Spring)*. 2011;**19**:1775-1783.
194. Stoeckel LE, Weller RE, Cook 3rd EW, Twieg DB, Knowlton RC, Cox JE. Widespread reward-system activation in obese women in response to pictures of high-calorie foods. *Neuroimage*. 2008;**41**:636-647.
195. Dimitropoulos AT, Tkach J, Ho A, Kennedy J. Greater corticolimbic activation to high-calorie food cues after eating in obese vs. normal-weight adults. *Appetite*. 2012;**58**(1):303-312.
196. Martin LE, Holsen LM, Chambers RJ, et al. Neural mechanisms associated with food motivation in obese and healthy weight adults. *Obesity (Silver Spring)*. 2010;**18**:254-260.
197. Stice E, Spoor S, Bohon C, Veldhuizen MG, Small DM. Relation of reward from food intake and anticipated food intake to obesity: a functional magnetic resonance imaging study. *J Abnorm Psychol*. 2008;**117**(4):924-935.
198. Connolly L, Coveleskie K, Kilpatrick LA, et al. Differences in brain responses between lean and obese women to a sweetened drink. *Neurogastroenterol Motil*. 2013;**25**(7):579-e460.
199. Bragulat V, Dzemidzic M, Bruno C, et al. Food-related odor probes of brain reward circuits during hunger: a pilot FMRI study. *Obesity (Silver Spring)*. 2010;**18**(8):1566-1571.
200. Haase L, Green E, Murphy C. Males and females show differential brain activation to taste when hungry and sated in gustatory and reward areas. *Appetite*. 2011;**57**(2):421-434.
201. Melasch J, Rullmann M, Hilbert A, et al. Sex differences in serotonin-hypothalamic connections underpin a diminished sense of emotional well-being with increasing body weight. *Int J Obes (Lond)*. 2016;**40**(8):1268-1277.
202. Cornier MA, Salzberg AK, Endly DC, Bessesen DH, Tregellas JR. Sex-based differences in the behavioral and neuronal responses to food. *Physiol Behav*. 2010;**99**(4):538-543.
203. Geliebter A, Pantazatos SP, McOuatt H, Puma L, Gibson CD, Atalayer D. Sex-based fMRI differences in obese humans in response to high vs. low energy food cues. *Behav Brain Res*. 2013;**243**:91-96.
204. Gupta A, Mayer EA, Hamadani K, et al. Sex differences in the influence of body mass index on anatomical architecture of brain networks. *Int J Obes (2005)*. 2017.
205. Gupta A, Mayer EA, Labus JS, et al. Sex commonalities and differences in obesity-related alterations in intrinsic brain activity and connectivity. *Obesity (Silver Spring)*. 2018;**26**(2):340-350.
206. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am J Psychiatry*. 2007;**164**(6):942-948.
207. Remes O, Brayne C, van der Linde R, Lafortune L. A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain Behav*. 2016;**6**(7):e00497.
208. Smink FRE, van Hoeken D, Dijkstra JK, Deen M, Oldehinkel AJ, Hoek HW. Self-esteem and peer-perceived social status in early adolescence and prediction of eating pathology in young adulthood. *Int J Eat Disord*. 2018;**51**(8):852-862.
209. Riedel BC, Thompson PM, Brinton RD. Age, APOE and sex: triad of risk of Alzheimer's disease. *J Steroid Biochem Mol Biol*. 2016;**160**:134-147.
210. Cahill L. Why sex matters for neuroscience. *Nat Rev Neurosci*. 2006;7(6):477-484.
211. Corre C, Friedel M, Vousden DA, et al. Separate effects of sex hormones and sex chromosomes on brain structure and function revealed by high-resolution magnetic resonance imaging and spatial navigation assessment of the Four Core Genotype mouse model. *Brain Struct Funct*. 2016;**221**(2):997-1016.
212. Mankiw C, Park MTM, Reardon PK, et al. Allometric analysis detects brain size-independent effects of sex and sex chromosome complement on human cerebellar organization. *J Neurosci*. 2017;**37**(21):5221-5231.
213. Joel D, Berman Z, Tavor I, et al. Sex beyond the genitalia: the human brain mosaic. *Proc Natl Acad Sci U S A*. 2015;**112**(50):15468-15473.
214. Wierenga LM, Bos MGN, Schreuders E, et al. Unraveling age, puberty and testosterone effects on subcortical brain development across adolescence. *Psychoneuroendocrinology*. 2018;**91**:105-114.
215. Mueller SC, Landré L, Wierckx K, T'Sjoen G. A structural magnetic resonance imaging study in transgender persons on cross-sex hormone therapy. *Neuroendocrinology*. 2017;**105**(2):123-130.
216. Merke DP, Fields JD, Keil MF, Vaituzis AC, Chrousos GP, Giedd JN. Children with classic congenital adrenal hyperplasia have decreased amygdala volume: potential prenatal and postnatal hormonal effects. *J Clin Endocrinol Metab*. 2003;**88**(4):1760-1765.
217. Martin CR, Osadchiy V, Kalani A, Mayer EA. The brain-gut-microbiome axis. *Cell Mol Gastroenterol Hepatol*. 2018;**6**(2):133-148.
218. Hagiwara SI, Kaushal E, Paruthiyil S, Pasricha PJ, Hasdemir B, Bhargava A. Gastric corticotropin-releasing factor influences mast cell infiltration in a rat model of functional dyspepsia. *Plos One*. 2018;**13**(9):e0203704.
219. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H -U. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *Int J Methods Psychiatr Res*. 2012;**21**(3):169-184.
220. Tesic A, Rodgers S, Müller M, et al. Sex differences in neurodevelopmental and common mental disorders examined from three epidemiological perspectives. *Psychiatry Res*. 2019;**278**:213-217.
221. Pinares-Garcia P, Stratikopoulos M, Zagato A, Loke H, Lee J. Sex: a significant risk factor for neurodevelopmental and neurodegenerative disorders. *Brain Sci*. 2018;8:154.
222. Kokras N, Hodes GE, Bangasser DA, Dalla C. Sex differences in the hypothalamic–pituitary–adrenal axis: an obstacle to antidepressant drug development? *Br J Pharmacol*. 2019;176 (21):4090-4106.

223. Cahill L, Aswad D. Sex influences on the brain: an issue whose time has come. *Neuron*. 2015;88(6):1084-1085.
224. Mendrek A, Mancini-Marïe A. Sex/gender differences in the brain and cognition in schizophrenia. *Neurosci Biobehav Rev*. 2016;67:57-78.
225. Bray GA, Heisel WE, Afshin A, et al. The science of obesity management: an Endocrine Society scientific statement. *Endocr Rev*. 2018;39(2):79-132.
226. Gartner DR, Taber DR, Hirsch JA, Robinson WR. The spatial distribution of gender differences in obesity prevalence differs from overall obesity prevalence among US adults. *Ann Epidemiol*. 2016;26(4):293-298.
227. Hales CM, Fryar CD, Carroll MD, Freedman DS, Ogden CL. Trends in obesity and severe obesity prevalence in US youth and adults by sex and age, 2007–2008 to 2015–2016. *JAMA*. 2018;319:1723-1725.
228. Mond JM, Baune BT. Overweight, medical comorbidity and health-related quality of life in a community sample of women and men. *Obesity (Silver Spring)*. 2009;17(8):1627-1634.
229. Lovejoy JC, Sainsbury A; Stock Conference 2008 Working Group. Sex differences in obesity and the regulation of energy homeostasis. *Obes Rev*. 2009;10(2):154-167.
230. Kelly T, Yang W, Chen CS, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. *Int J Obes (Lond)*. 2008;32:1431-1437.
231. Power ML, Schulkin J. Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. *Br J Nutr*. 2008;99(5):931-940.
232. Weltens N, Iven J, Van Oudenhove L, Kano M. The gut-brain axis in health neuroscience: implications for functional gastrointestinal disorders and appetite regulation. *Ann N Y Acad Sci*. 2018;1428(1):129-150.
233. Leigh SJ, Morris MJ. Diet, inflammation and the gut microbiome: Mechanisms for obesity-associated cognitive impairment. *Biochim Biophys Acta Mol Basis Dis*. 2020;1866(6):165767.
234. Torres-Fuentes C, Schellekens H, Dinan TG, Cryan JF. The microbiota-gut-brain axis in obesity. *Lancet Gastroenterol Hepatol*. 2017;2(10):747-756.
235. Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. *Curr Opin Gastroenterol*. 2015;31(1):69-75.
236. Osadchiy V, Martin CR, Mayer EA. The gut-brain axis and the microbiome: mechanisms and clinical implications. *Clin Gastroenterol Hepatol*. 2019;17(2):322-332.
237. Dong TS, Gupta A. Influence of early life, diet, and the environment on the microbiome. *Clin Gastroenterol Hepatol*. 2019;17(2):231-242.
238. Jašarević E, Morrison KE, Bale TL. Sex differences in the gut microbiome-brain axis across the lifespan. *Philos Trans R Soc Lond B Biol Sci*. 2016;371(1688):20150122.
239. Markle JG, Frank DN, Mortin-Toth S, et al. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science*. 2013;339(6123):1084-1088.
240. Yurkovetskiy L, Burrows M, Khan AA, et al. Gender bias in autoimmunity is influenced by microbiota. *Immunity*. 2013;39(2):400-412.
241. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222-227.
242. Collins SM, Kassam Z, Bercik P. The adoptive transfer of behavioral phenotype via the intestinal microbiota: experimental evidence and clinical implications. *Curr Opin Microbiol*. 2013;16(3):240-245.
243. Chen KL, Madak-Erdogan Z. Estrogen and microbiota crosstalk: should we pay attention? *Trends Endocrinol Metab*. 2016;27(11):752-755.
244. Plottel CS, Blaser MJ. Microbiome and malignancy. *Cell Host Microbe*. 2011;10(4):324-335.
245. Moser VA, Pike CJ. Obesity and sex interact in the regulation of Alzheimer's disease. *Neurosci Biobehav Rev*. 2016;67:102-118.
246. Stachowiak G, Pertyński T, Pertyńska-Marczewska M. Metabolic disorders in menopause. *Prz Menopauzalny*. 2015;14(1):59-64.
247. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *Jama*. 1998;280(7):605-613.
248. Lichanska AM, Waters MJ. How growth hormone controls growth, obesity and sexual dimorphism. *Trends Genet*. 2008;24(1):41-47.
249. Lee C, Kim J, Jung Y. Potential therapeutic application of estrogen in gender disparity of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *Cells*. 2019;8(10):1259.
250. Monteiro R, Teixeira D, Calhau C. Estrogen signaling in metabolic inflammation. *Mediators Inflamm*. 2014;2014:615917.
251. Eaton SA, Sethi JK. Immunometabolic links between estrogen, adipose tissue and female reproductive metabolism. *Biology (Basel)*. 2019;8(1):8.
252. Al-Suhaimi EA, Shehzad A. Leptin, resistin and visfatin: the missing link between endocrine metabolic disorders and immunity. *Eur J Med Res*. 2013;18:12.
253. O'Mahony SM, Clarke G, Borre YE, Dinan TG, Cryan JF. Serotonin, tryptophan metabolism and the brain-gut-microbiome axis. *Behav Brain Res*. 2015;277:32-48.
254. Mawe GM, Hoffman JM. Serotonin signalling in the gut–functions, dysfunctions and therapeutic targets. *Nat Rev Gastroenterol Hepatol*. 2013;10(8):473-486.
255. Kennedy PJ, Cryan JF, Dinan TG, Clarke G. Kynurenine pathway metabolism and the microbiota-gut-brain axis. *Neuropharmacology*. 2017;112(Pt B):399-412.
256. Mangge H, Summers KL, Meinitzer A, et al. Obesity-related dysregulation of the tryptophan-kynurenine metabolism: role of age and parameters of the metabolic syndrome. *Obesity (Silver Spring)*. 2014;22(1):195-201.
257. Favennec M, Hennart B, Caiazzo R, et al. The kynurenine pathway is activated in human obesity and shifted toward kynurenine monooxygenase activation. *Obesity (Silver Spring)*. 2015;23(10):2066-2074.
258. Cazettes F, Cohen JI, Yau PL, Talbot H, Convit A. Obesity-mediated inflammation may damage the brain circuit that regulates food intake. *Brain Res*. 2011;1373:101-109.
259. Castanon N, Lasselin J, Capuron L. Neuropsychiatric comorbidity in obesity: role of inflammatory processes. *Front Endocrinol (Lausanne)*. 2014;5:74.

260. Jhamandas K, Boegman RJ, Beninger RJ, Bialik M. Quinolinate-induced cortical cholinergic damage: modulation by tryptophan metabolites. *Brain Res.* 1990;**529**(1-2):185-191.
261. Raheja UK, Fuchs D, Giegling I, et al. In psychiatrically healthy individuals, overweight women but not men have lower tryptophan levels. *Pteridines.* 2015;**26**(2):79-84.
262. Theofylaktopoulou D, Midttun Ø, Ulvik A, et al. A community-based study on determinants of circulating markers of cellular immune activation and kynurenines: the Hordaland Health Study. *Clin Exp Immunol.* 2013;**173**(1):121-130.
263. Cortese S, Faraone SV, Bernardi S, Wang S, Blanco C. Gender differences in adult attention-deficit/hyperactivity disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *J Clin Psychiatry.* 2016;**77**(4):e421-e428.
264. Wittchen HU, Jacobi F, Rehm J, et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol.* 2011;**21**:655-679.
265. Salk RH, Hyde JS, Abramson LY. Gender differences in depression in representative national samples: meta-analyses of diagnoses and symptoms. *Psychol Bull.* 2017;**143**(8):783-822.
266. Diflorio A, Jones I. Is sex important? Gender differences in bipolar disorder. *Int Rev Psychiatry.* 2010;**22**(5):437-452.
267. Sramek JJ, Murphy MF, Cutler NR. Sex differences in the psychopharmacological treatment of depression. *Dialogues Clin Neurosci.* 2016;**18**(4):447-457.
268. Kokras N, Dalla C. Preclinical sex differences in depression and antidepressant response: implications for clinical research. *J Neurosci Res.* 2017;**95**(1-2):731-736.
269. Bangasser DA. Sex differences in stress-related receptors: "micro" differences with "macro" implications for mood and anxiety disorders. *Biol Sex Differ.* 2013;**4**(1):2.
270. Bredewold R, Schiavo JK, van der Hart M, Verreij M, Veenema AH. Dynamic changes in extracellular release of GABA and glutamate in the lateral septum during social play behavior in juvenile rats: implications for sex-specific regulation of social play behavior. *Neuroscience.* 2015;**307**:117-127.
271. Campi KL, Greenberg GD, Kapoor A, Ziegler TE, Trainor BC. Sex differences in effects of dopamine D1 receptors on social withdrawal. *Neuropharmacology.* 2014;**77**:208-216.
272. Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE. Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov.* 2017;**16**(12):829-842.
273. Bangasser DA, Wiersielis KR. Sex differences in stress responses: a critical role for corticotropin-releasing factor. *Hormones (Athens).* 2018;**17**(1):5-13.
274. Vuppaladhadiam L, Ehsan C, Akkati M, Bhargava A. Corticotropin-releasing factor family: a stress hormone-receptor system's emerging role in mediating sex-specific signaling. *Cells.* 2020;**9**(4):839
275. Pal K, Swaminathan K, Xu HE, Pioszak AA. Structural basis for hormone recognition by the Human CRFR2{α} G protein-coupled receptor. *J Biol Chem.* 2010;**285**(51):40351-40361.
276. Iwasaki-Sekino A, Mano-Otagiri A, Ohata H, Yamauchi N, Shibasaki T. Gender differences in corticotropin and corticosterone secretion and corticotropin-releasing factor mRNA expression in the paraventricular nucleus of the hypothalamus and the central nucleus of the amygdala in response to footshock stress or psychological stress in rats. *Psychoneuroendocrinology.* 2009;**34**(2):226-237.
277. Viau V, Bingham B, Davis J, Lee P, Wong M. Gender and puberty interact on the stress-induced activation of parvocellular neurosecretory neurons and corticotropin-releasing hormone messenger ribonucleic acid expression in the rat. *Endocrinology.* 2005;**146**(1):137-146.
278. Stinnett GS, Westphal NJ, Seasholtz AF. Pituitary CRH-binding protein and stress in female mice. *Physiol Behav.* 2015;**150**:16-23.
279. Gallucci WT, Baum A, Laue L, et al. Sex differences in sensitivity of the hypothalamic-pituitary-adrenal axis. *Health Psychol.* 1993;**12**(5):420-425.
280. Valentino RJ, Van Bockstaele EJ. Corticotropin-releasing factor: putative neurotransmitter actions of a neurohormone. In: D Pfaff AA, Etgen A, Fahrbach S, Moss R, Rubin R, eds. *Hormones, Brain and Behavior.* Vol. 4. San Diego: Academic Press; 2002:81-102.
281. Weathington JM, Cooke BM. Corticotropin-releasing factor receptor binding in the amygdala changes across puberty in a sex-specific manner. *Endocrinology.* 2012;**153**(12):5701-5705.
282. Crestani CC, Alves FH, Gomes FV, Resstel LB, Correa FM, Herman JP. Mechanisms in the bed nucleus of the stria terminalis involved in control of autonomic and neuroendocrine functions: a review. *Curr Neuropharmacol.* 2013;**11**(2):141-159.
283. Lim MM, Nair HP, Young LJ. Species and sex differences in brain distribution of corticotropin-releasing factor receptor subtypes 1 and 2 in monogamous and promiscuous vole species. *J Comp Neurol.* 2005;**487**(1):75-92.
284. Catalano RD, Kyriakou T, Chen J, Easton A, Hillhouse EW. Regulation of corticotropin-releasing hormone type 2 receptors by multiple promoters and alternative splicing: identification of multiple splice variants. *Mol Endocrinol.* 2003;**17**(3):395-410.
285. Weiser MJ, Goel N, Sandau US, Bale TL, Handa RJ. Androgen regulation of corticotropin-releasing hormone receptor 2 (CRHR2) mRNA expression and receptor binding in the rat brain. *Exp Neurol.* 2008;**214**(1):62-68.
286. Aston-Jones G, Bloom FE. Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep-waking cycle. *J Neurosci.* 1981;**1**(8):876-886.
287. Aston-Jones G. *CHAPTER 11 - Locus coeruleus, A5 and A7 noradrenergic cell groups A2 - Paxinos, George. The Rat Nervous System (THIRD EDITION).* Burlington: Academic Press; 2004:259-294.
288. Berridge CW, Page ME, Valentino RJ, Foote SL. Effects of locus coeruleus inactivation on electroencephalographic activity in neocortex and hippocampus. *Neuroscience.* 1993;**55**(2):381-393.
289. Berridge CW, Abercrombie ED. Relationship between locus coeruleus discharge rates and rates of norepinephrine release within neocortex as assessed by in vivo microdialysis. *Neuroscience.* 1999;**93**(4):1263-1270.
290. Curtis AL, Lechner SM, Pavcovich LA, Valentino RJ. Activation of the locus coeruleus noradrenergic system by intracoerulear microinfusion of corticotropin-releasing factor: effects on

discharge rate, cortical norepinephrine levels and cortical electroencephalographic activity. *J Pharmacol Exp Ther.* 1997;**281**(1):163-172.
291. Valentino RJ, Curtis AL, Page ME, Pavcovich LA, Florin-Lechner SM. Activation of the locus ceruleus brain noradrenergic system during stress: circuitry, consequences, and regulation. *Adv Pharmacol.* 1998;**42**:781-784.
292. Gold PW, Chrousos GP. Organization of the stress system and its dysregulation in melancholic and atypical depression: high vs low CRH/NE states. *Mol Psychiatry.* 2002;**7**(3):254-275.
293. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5.* 5th ed. American Psychiatric Publishing; 2013.
294. Hagiwara SI, Hasdemir B, Heyman MB, Chang L, Bhargava A. Plasma corticotropin-releasing factor receptors and B7-2(+) extracellular vesicles in blood correlate with irritable bowel syndrome disease severity. *Cells.* 2019;**8**(2):101.
295. Grammatopoulos DK, Randeva HS, Levine MA, Kanellopoulou KA, Hillhouse EW. Rat cerebral cortex corticotropin-releasing hormone receptors: evidence for receptor coupling to multiple G-proteins. *J Neurochem.* 2001;**76**(2):509-519.
296. Chen FM, Bilezikjian LM, Perrin MH, Rivier J, Vale W. Corticotropin releasing factor receptor-mediated stimulation of adenylate cyclase activity in the rat brain. *Brain Res.* 1986;**381**(1):49-57.
297. De Souza EB. Corticotropin-releasing factor receptors: physiology, pharmacology, biochemistry and role in central nervous system and immune disorders. *Psychoneuroendocrinology.* 1995;**20**(8):789-819.
298. Bangasser DA, Curtis A, Reyes BA, et al. Sex differences in corticotropin-releasing factor receptor signaling and trafficking: potential role in female vulnerability to stress-related psychopathology. *Mol Psychiatry.* 2010;**15**:877, 896-904.
299. Valentino RJ, Van Bockstaele E, Bangasser D. Sex-specific cell signaling: the corticotropin-releasing factor receptor model. *Trends Pharmacol Sci.* 2013;**34**(8):437-444.
300. Bangasser DA, Dong H, Carroll J, et al. Corticotropin-releasing factor overexpression gives rise to sex differences in Alzheimer's disease-related signaling. *Mol Psychiatry.* 2017;**22**(8):1126-1133.
301. Komuro H, Sato N, Sasaki A, et al. Corticotropin-releasing hormone receptor 2 gene variants in irritable bowel syndrome. *PLoS One.* 2016;**11**(1):e0147817.
302. Wolf EJ, Mitchell KS, Logue MW, et al. Corticotropin releasing hormone receptor 2 (CRHR-2) gene is associated with decreased risk and severity of posttraumatic stress disorder in women. *Depress Anxiety.* 2013;**30**(12):1161-1169.
303. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet.* 2018;**50**(11):1505-1513.
304. Kautzky-Willer A, Harreiter J, Pacini G. Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus. *Endocr Rev.* 2016;**37**(3):278-316.
305. Bergmann N, Gyntelberg F, Faber J. The appraisal of chronic stress and the development of the metabolic syndrome: a systematic review of prospective cohort studies. *Endocr Connect.* 2014;**3**(2):R55-R80.
306. Fuller-Jackson JP, Dordevic AL, Clarke IJ, Henry BA. Effect of sex and sex steroids on brown adipose tissue heat production in humans. *Eur J Endocrinol.* 2020;**183**(3):343-355.
307. Romagnuolo J, Talluri J, Kennard E, et al. Clinical profile, etiology, and treatment of chronic pancreatitis in North American Women: analysis of a large multicenter cohort. *Pancreas.* 2016;**45**(7):934-940.
308. Violin JD, Lefkowitz RJ. Beta-arrestin-biased ligands at seven-transmembrane receptors. *Trends Pharmacol Sci.* 2007;**28**(8):416-422.
309. Whalen EJ, Rajagopal S, Lefkowitz RJ. Therapeutic potential of β-arrestin- and G protein-biased agonists. *Trends Mol Med.* 2011;**17**(3):126-139.
310. Murrough JW, Charney DS. Corticotropin-releasing factor type 1 receptor antagonists for stress-related disorders: time to call it quits? *Biol Psychiatry.* 2017;**82**(12):858-860.
311. Hasdemir B, Mahajan S, Oses-Prieto J, et al. Actin cytoskeleton-dependent regulation of corticotropin-releasing factor receptor heteromers. *Mol Biol Cell.* 2017;**28**(18):2386-2399.
312. Mikhailova MV, Mayeux PR, Jurkevich A, et al. Heterooligomerization between vasotocin and corticotropin-releasing hormone (CRH) receptors augments CRH-stimulated 3',5'-cyclic adenosine monophosphate production. *Mol Endocrinol.* 2007;**21**(9):2178-2188.
313. Murat B, Devost D, Andrés M, et al. V1b and CRHR1 receptor heterodimerization mediates synergistic biological actions of vasopressin and CRH. *Mol Endocrinol.* 2012;**26**(3):502-520.
314. Chaki S, Nakazato A, Kennis L, et al. Anxiolytic- and antidepressant-like profile of a new CRF1 receptor antagonist, R278995/CRA0450. *Eur J Pharmacol.* 2004;**485**(1-3):145-158.
315. Deak T, Nguyen KT, Ehrlich AL, et al. The impact of the nonpeptide corticotropin-releasing hormone antagonist antalarmin on behavioral and endocrine responses to stress. *Endocrinology.* 1999;**140**(1):79-86.
316. Mansbach RS, Brooks EN, Chen YL. Antidepressant-like effects of CP-154,526, a selective CRF1 receptor antagonist. *Eur J Pharmacol.* 1997;**323**(1):21-26.
317. Zorrilla EP, Valdez GR, Nozulak J, Koob GF, Markou A. Effects of antalarmin, a CRF type 1 receptor antagonist, on anxiety-like behavior and motor activation in the rat. *Brain Res.* 2002;**952**(2):188-199.
318. Howerton AR, Roland AV, Fluharty JM, et al. Sex differences in corticotropin-releasing factor receptor-1 action within the dorsal raphe nucleus in stress responsivity. *Biol Psychiatry.* 2014;**75**(11):873-883.
319. Ising M, Zimmermann US, Künzel HE, et al. High-affinity CRF1 receptor antagonist NBI-34041: preclinical and clinical data suggest safety and efficacy in attenuating elevated stress response. *Neuropsychopharmacology.* 2007;**32**(9):1941-1949.
320. Beery AK, Zucker I. Sex bias in neuroscience and biomedical research. *Neurosci Biobehav Rev.* 2011;**35**(3):565-572.
321. Will TR, Proaño SB, Thomas AM, et al. Problems and progress regarding sex bias and omission in neuroscience research. *eNeuro.* 2017;**4**:ENEURO.0278-0217.2017.
322. Rademaker M. Do women have more adverse drug reactions? *Am J Clin Dermatol.* 2001;**2**:349-351.

323. McKenzie BE. Guidelines and requirements for the evaluation of contraceptive steroids. *Toxicol Pathol*. 1989;**17**(2):377-384.
324. Farkouh A, Riedl T, Gottardi R, Czejka M, Kautzky-Willer A. Sex-related differences in pharmacokinetics and pharmacodynamics of frequently prescribed drugs: a review of the literature. *Adv Ther*. 2020;**37**(2):644-655.
325. Franconi F, Campesi I. Pharmacogenomics, pharmacokinetics and pharmacodynamics: interaction with biological differences between men and women. *Br J Pharmacol*. 2014;**171**(3):580-594.
326. Gartlehner G, Chapman A, Strobelberger M, Thaler K. Differences in efficacy and safety of pharmaceutical treatments between men and women: an umbrella review. *Plos One*. 2010;**5**(7):e11895.
327. Viramontes BE, Camilleri M, McKinzie S, Pardi DS, Burton D, Thomforde GM. Gender-related differences in slowing colonic transit by a 5-HT3 antagonist in subjects with diarrhea-predominant irritable bowel syndrome. *Am J Gastroenterol*. 2001;**96**(9):2671-2676.
328. Mosca L, Barrett-Connor E, Wenger NK. Sex/gender differences in cardiovascular disease prevention: what a difference a decade makes. *Circulation*. 2011;**124**(19):2145-2154.
329. WHO. *Cardiovascular Disease*. 2017. Accessed April 2020. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
330. Maas AH, Appelman YE. Gender differences in coronary heart disease. *Neth Heart J*. 2010;**18**(12):598-602.
331. Regitz-Zagrosek V. Unsettled issues and future directions for research on cardiovascular diseases in women. *Korean Circ J*. 2018;**48**(9):792-812.
332. Colafella KMM, Denton KM. Sex-specific differences in hypertension and associated cardiovascular disease. *Nat Rev Nephrol*. 2018;**14**(3):185-201.
333. Mannon EC, Ray SC, Ryan MJ, Sullivan JC. Does sex matter?: an update on the implementation of sex as a biological variable in research. *Am J Physiol Renal Physiol*. 2020;**318**(2):F329-F331.
334. Guyton AC, Coleman TG, Young DB, Lohmeier TE, DeClue JW. Salt balance and long-term blood pressure control. *Annu Rev Med*. 1980;**31**:15-27.
335. Davel AP, Jaffe IZ, Tostes RC, Jaisser F, Belin de Chantemèle EJ. New roles of aldosterone and mineralocorticoid receptors in cardiovascular disease: translational and sex-specific effects. *Am J Physiol Heart Circ Physiol*. 2018;**315**(4):H989-H999.
336. Wiinberg N, Høegholm A, Christensen HR, et al. 24-h ambulatory blood pressure in 352 normal Danish subjects, related to age and gender. *Am J Hypertens*. 1995;**8**(10 Pt 1):978-986.
337. Yong LC, Kuller LH, Rutan G, Bunker C. Longitudinal study of blood pressure: changes and determinants from adolescence to middle age. The Dormont High School follow-up study, 1957–1963 to 1989–1990. *Am J Epidemiol*. 1993;**138**:973-983.
338. Sandberg K, Ji H. Sex differences in primary hypertension. *Biol Sex Differ*. 2012;**3**(1):7.
339. Himmelmann A, Svensson A, Hansson L. Influence of sex on blood pressure and left ventricular mass in adolescents: the hypertension in pregnancy offspring study. *J Hum Hypertens*. 1994;**8**(7):485-490.
340. Stamler J, Stamler R, Riedlinger WF, Algera G, Roberts RH. Hypertension screening of 1 million Americans. Community Hypertension Evaluation Clinic (CHEC) program, 1973 through 1975. *JAMA*. 1976;**235**:2299-2306.
341. Burt V, Whelton P, Roccella E, et al. Prevalence of hypertension in the US adult population: results from the Third National Health and Nutrition Examination Survey, 1988–1991. *Am Heart Assoc*. 1995;**25**:305-313.
342. Benjamin EJ, Blaha MJ, Chiuve SE, et al.; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics-2017 update: a report from the American Heart Association. *Circulation*. 2017;**135**(10):e146-e603.
343. Engberding N, Wenger NK. Management of hypertension in women. *Hypertens Res*. 2012;**35**(3):251-260.
344. Roger VL, Go AS, Lloyd-Jones DM, et al.; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Executive summary: heart disease and stroke statistics–2012 update: a report from the American Heart Association. *Circulation*. 2012;**125**(1):188-197.
345. Hilliard LM, Sampson AK, Brown RD, Denton KM. The "his and hers" of the renin-angiotensin system. *Curr Hypertens Rep*. 2013;**15**(1):71-79.
346. Sampson AK, Hilliard LM, Moritz KM, et al. The arterial depressor response to chronic low-dose angiotensin II infusion in female rats is estrogen dependent. *Am J Physiol Regul Integr Comp Physiol*. 2012;**302**(1):R159-R165.
347. Miller JA, Anacta LA, Cattran DC. Impact of gender on the renal response to angiotensin II. *Kidney Int*. 1999;**55**(1):278-285.
348. Colafella KM, Hilliard LM, Denton KM. Epochs in the depressor/pressor balance of the renin-angiotensin system. *Clin Sci (Lond)*. 2016;**130**(10):761-771.
349. Phoon S, Howes LG. Forearm vasodilator response to angiotensin II in elderly women receiving candesartan: role of AT(2)- receptors. *J Renin Angiotensin Aldosterone Syst*. 2002;**3**(1):36-39.
350. Mirabito KM, Hilliard LM, Kett MM, et al. Sex- and age-related differences in the chronic pressure-natriuresis relationship: role of the angiotensin type 2 receptor. *Am J Physiol Renal Physiol*. 2014;**307**(8):F901-F907.
351. Ichiki T, Kambayashi Y, Inagami T. Molecular cloning and expression of angiotensin II type 2 receptor gene. *Adv Exp Med Biol*. 1996;**396**:145-152.
352. Xue Q, Xiao D, Zhang L. Estrogen regulates angiotensin II receptor expression patterns and protects the heart from ischemic injury in female rats. *Biol Reprod*. 2015;**93**(1):6.
353. Sanderink GJ, Artur Y, Schiele F, Gueguen R, Siest G. Alanine aminopeptidase in serum: biological variations and reference limits. *Clin Chem*. 1988;**34**(7):1422-1426.
354. Mueller PW, Phillips DL, Steinberg KK. Alanine aminopeptidase in serum: automated optimized assay, and effects of age, sex, smoking, and alcohol consumption in a selected population. *Clin Chem*. 1987;**33**(3):363-366.
355. Mizutani S, Yamada R, Kurauchi O, Ito Y, Narita O, Tomoda Y. Serum aminopeptidase A (AAP) in normal pregnancy and pregnancy complicated by pre-eclampsia. *Arch Gynecol*. 1987;**240**(1):27-31.
356. Hariyama Y, Itakura A, Okamura M, et al. Placental aminopeptidase A as a possible barrier of angiotensin II between mother and fetus. *Placenta*. 2000;**21**(7):621-627.

357. Gohar EY, Pollock DM. Sex-specific contributions of endothelin to hypertension. *Curr Hypertens Rep*. 2018;**20**(7):58.
358. Tamma G, Goswami N, Reichmuth J, De Santo NG, Valenti G. Aquaporins, vasopressin, and aging: current perspectives. *Endocrinology*. 2015;**156**(3):777-788.
359. Juul KV, Bichet DG, Nielsen S, Nørgaard JP. The physiological and pathophysiological functions of renal and extrarenal vasopressin V2 receptors. *Am J Physiol Renal Physiol*. 2014;**306**(9):F931-F940.
360. Clemmer JS, Faulkner JL, Mullen AJ, Butler KR, Hester RL. Sex-specific responses to mineralocorticoid receptor antagonism in hypertensive African American males and females. *Biol Sex Differ*. 2019;**10**(1):24.
361. Faulkner JL, Kennard S, Huby AC, et al. Progesterone predisposes females to obesity-associated leptin-mediated endothelial dysfunction via upregulating endothelial MR (Mineralocorticoid Receptor) expression. *Hypertension*. 2019;**74**(3):678-686.
362. Veiras LC, Girardi ACC, Curry J, et al. Sexual dimorphic pattern of renal transporters and electrolyte homeostasis. *J Am Soc Nephrol*. 2017;**28**(12):3504-3517.
363. Mirabito Colafella KM, Samuel CS, Denton KM. Relaxin contributes to the regulation of arterial pressure in adult female mice. *Clin Sci (Lond)*. 2017;**131**(23):2795-2805.
364. Danielson LA, Kercher LJ, Conrad KP. Impact of gender and endothelin on renal vasodilation and hyperfiltration induced by relaxin in conscious rats. *Am J Physiol Regul Integr Comp Physiol*. 2000;**279**(4):R1298-R1304.
365. Chapman AB, Abraham WT, Zamudio S, et al. Temporal relationships between hormonal and hemodynamic changes in early human pregnancy. *Kidney Int*. 1998;**54**(6):2056-2063.
366. Anton L, Merrill DC, Neves LA, et al. The uterine placental bed Renin-Angiotensin system in normal and preeclamptic pregnancy. *Endocrinology*. 2009;**150**(9):4316-4325.
367. Brouwers L, van der Meiden-van Roest AJ, Savelkoul C, et al. Recurrence of pre-eclampsia and the risk of future hypertension and cardiovascular disease: a systematic review and meta-analysis. *BJOG*. 2018;**125**(13):1642-1654.
368. Moreau KL, Babcock MC, Hildreth KL. Sex differences in vascular aging in response to testosterone. *Biol Sex Differ*. 2020;**11**(1):18.
369. Reckelhoff JF. Androgens and blood pressure control: sex differences and mechanisms. *Mayo Clin Proc*. 2019;**94**(3):536-543.
370. Faulkner JL, Belin de Chantemèle EJ. Sex hormones, aging and cardiometabolic syndrome. *Biol Sex Differ*. 2019;**10**(1):30.
371. Sabbatini AR, Kararigas G. Estrogen-related mechanisms in sex differences of hypertension and target organ damage. *Biol Sex Differ*. 2020;**11**(1):31.
372. Oelkers WK. Effects of estrogens and progestogens on the renin-aldosterone system and blood pressure. *Steroids*. 1996;**61**(4):166-171.
373. Cobin RH, Goodman NF; AACE Reproductive Endocrinology Scientific Committee. American Association of Clinical Endocrinologists and American College of Endocrinology position statement on menopause-2017 update. *Endocr Pract*. 2017;**23**(7):869-880.
374. Gencer B, Mach F. Testosterone: a hormone preventing cardiovascular disease or a therapy increasing cardiovascular events? *Eur Heart J*. 2016;**37**(48):3569-3575.
375. Fuchs MM, Attenhofer Jost C, Babovic-Vuksanovic D, Connolly HM, Egbe A. Long-term outcomes in patients with turner syndrome: a 68-year follow-up. *J Am Heart Assoc*. 2019;**8**(11):e011501.
376. Price WH, Clayton JF, Collyer S, De Mey R, Wilson J. Mortality ratios, life expectancy, and causes of death in patients with Turner's syndrome. *J Epidemiol Community Health*. 1986;**40**(2):97-102.
377. Schoemaker MJ, Swerdlow AJ, Higgins CD, Wright AF, Jacobs PA; United Kingdom Clinical Cytogenetics Group. Mortality in women with Turner syndrome in Great Britain: a national cohort study. *J Clin Endocrinol Metab*. 2008;**93**(12):4735-4742.
378. Gravholt CH, Viuff MH, Brun S, Stochholm K, Andersen NH. Turner syndrome: mechanisms and management. *Nat Rev Endocrinol*. 2019;**15**(10):601-614.
379. Accardo G, Amoresano Paglionico V, Di Fraia R, et al. Management of cardiovascular complications in Klinefelter syndrome patients. *Expert Rev Endocrinol Metab*. 2019;**14**(2):145-152.
380. Pasquali D, Arcopinto M, Renzullo A, et al. Cardiovascular abnormalities in Klinefelter syndrome. *Int J Cardiol*. 2013;**168**(2):754-759.
381. Bojesen A, Juul S, Birkebaek N, Gravholt CH. Increased mortality in Klinefelter syndrome. *J Clin Endocrinol Metab*. 2004;**89**(8):3830-3834.
382. Swerdlow AJ, Higgins CD, Schoemaker MJ, Wright AF, Jacobs PA; United Kingdom Clinical Cytogenetics Group. Mortality in patients with Klinefelter syndrome in Britain: a cohort study. *J Clin Endocrinol Metab*. 2005;**90**(12):6516-6522.
383. Huby RD, Glaves P, Jackson R. The incidence of sexually dimorphic gene expression varies greatly between tissues in the rat. *Plos One*. 2014;**9**(12):e115792.
384. Mayne BT, Bianco-Miotto T, Buckberry S, et al. Large scale gene expression meta-analysis reveals tissue-specific, sex-biased gene expression in humans. *Front Genet*. 2016;**7**:183.
385. Kwekel JC, Vijay V, Desai VG, Moland CL, Fuscoe JC. Age and sex differences in kidney microRNA expression during the life span of F344 rats. *Biol Sex Differ*. 2015;**6**(1):1.
386. Kwekel JC, Desai VG, Moland CL, Vijay V, Fuscoe JC. Life cycle analysis of kidney gene expression in male F344 rats. *Plos One*. 2013;**8**(10):e75305.
387. Pan WH, Yeh WT, Hwu CM, Ho LT. Undiagnosed diabetes mellitus in Taiwanese subjects with impaired fasting glycemia: impact of female sex, central obesity, and short stature. *Chin J Physiol*. 2001;**44**(1):44-51.
388. Olivarius Nde F, Vestbo E, Andreasen AH, Mogensen CE. Renal involvement is related to body height in newly diagnosed diabetic women aged 40 years or over. *Diabetes Metab*. 2001;**27**(1):14-18.
389. Neugarten J, Silbiger SR. The impact of gender on renal transplantation. *Transplantation*. 1994;**58**(11):1145-1152.
390. Ishikawa I, Maeda K, Nakai S, Kawaguchi Y. Gender difference in the mean age at the induction of hemodialysis in patients with autosomal dominant polycystic kidney disease. *Am J Kidney Dis*. 2000;**35**(6):1072-1075.
391. Hannedouche T, Chauveau P, Kalou F, Albouze G, Lacour B, Jungers P. Factors affecting progression in advanced chronic renal failure. *Clin Nephrol*. 1993;**39**(6):312-320.

392. Coggins CH, Breyer Lewis J, Caggiula AW, Castaldo LS, Klahr S, Wang SR. Differences between women and men with chronic renal disease. *Nephrol Dial Transplant.* 1998;**13**(6):1430-1437.
393. Beale AL, Nanayakkara S, Segan L, et al. Sex differences in heart failure with preserved ejection fraction pathophysiology: a detailed invasive hemodynamic and echocardiographic analysis. *JACC Heart Fail.* 2019;**7**(3):239-249.
394. Turnbull F, Woodward M, Neal B, et al.; Blood Pressure Lowering Treatment Trialists' Collaboration. Do men and women respond differently to blood pressure-lowering treatment? Results of prospectively designed overviews of randomized trials. *Eur Heart J.* 2008;**29**(21):2669-2680.
395. Tadic M, Cuspidi C, Grassi G, Ivanovic B. Gender-specific therapeutic approach in arterial hypertension - challenges ahead. *Pharmacol Res.* 2019;**141**:181-188.
396. McDonough AA, Nguyen MT. Maintaining balance under pressure: integrated regulation of renal transporters during hypertension. *Hypertension.* 2015;**66**(3):450-455.
397. Kasai F, Hirayama N, Ozawa M, Iemura M, Kohara A. Changes of heterogeneous cell POPULATIONS in the Ishikawa cell line during long-term culture: proposal for an in vitro clonal evolution model of tumor cells. *Genomics.* 2016;**107**(6):259-266.
398. Yang DP, Rosanoff EI. Specific chromosome changes associated with rabbit cell lines cultured in vitro. *Cytogenet Cell Genet.* 1977;**18**(4):212-230.
399. Gebhard C, Regitz-Zagrosek V, Neuhauser HK, Morgan R, Klein SL. Impact of sex and gender on COVID-19 outcomes in Europe. *Biol Sex Differ.* 2020;**11**(1):29.
400. Jin JM, Bai P, He W, et al. Gender differences in patients with COVID-19: focus on severity and mortality. *Front Public Health.* 2020;8:152.
401. Kragholm K, Andersen MP, Gerds TA, et al. Association between male sex and outcomes of Coronavirus Disease 2019 (Covid-19) - a Danish nationwide, register-based study. *Clin Infect Dis.* Published online ahead of print July 8, 2020. doi:10.1093/cid/ciaa924
402. Kassotis CD, Vandenberg LN, Demeneix BA, Porta M, Slama R, Trasande L. Endocrine-disrupting chemicals: economic, regulatory, and policy implications. *Lancet Diabetes Endocrinol.* 2020;**8**(8):719-730.
403. Waxman DJ, Holloway MG. Sex differences in the expression of hepatic drug metabolizing enzymes. *Mol Pharmacol.* 2009;**76**(2):215-228.
404. Brehm E, Flaws JA. Transgenerational effects of endocrine-disrupting chemicals on male and female reproduction. *Endocrinology.* 2019;**160**(6):1421-1435.
405. De Vries GJ. Minireview: sex differences in adult and developing brains: compensation, compensation, compensation. *Endocrinology.* 2004;**145**(3):1063-1068.
406. Ichise H, Hori A, Shiozawa S, et al. Establishment of a tamoxifen-inducible Cre-driver mouse strain for widespread and temporal genetic modification in adult mice. *Exp Anim.* 2016;**65**(3):231-244.
407. Ye R, Wang QA, Tao C, et al. Impact of tamoxifen on adipocyte lineage tracing: Inducer of adipogenesis and prolonged nuclear translocation of Cre recombinase. *Mol Metab.* 2015;**4**(11):771-778.
408. Levy N, Paruthiyil S, Zhao X, et al. Unliganded estrogen receptor-beta regulation of genes is inhibited by tamoxifen. *Mol Cell Endocrinol.* 2010;**315**(1-2):201-207.

International Journal of Transgenderism

Taylor & Francis

ISSN: 1553-2739 (Print) 1434-4599 (Online) Journal homepage: http://www.tandfonline.com/loi/wijt20

A critical commentary on follow-up studies and "desistance" theories about transgender and gender-nonconforming children

Julia Temple Newhook, Jake Pyne, Kelley Winters, Stephen Feder, Cindy Holmes, Jemma Tosh, Mari-Lynne Sinnott, Ally Jamieson & Sarah Pickett

To cite this article: Julia Temple Newhook, Jake Pyne, Kelley Winters, Stephen Feder, Cindy Holmes, Jemma Tosh, Mari-Lynne Sinnott, Ally Jamieson & Sarah Pickett (2018): A critical commentary on follow-up studies and "desistance" theories about transgender and gender-nonconforming children, International Journal of Transgenderism, DOI: 10.1080/15532739.2018.1456390

To link to this article: https://doi.org/10.1080/15532739.2018.1456390

Published online: 26 Apr 2018.

INTERNATIONAL JOURNAL OF TRANSGENDERISM
https://doi.org/10.1080/15532739.2018.1456390

Taylor & Francis
Taylor & Francis Group

Check for updates

A critical commentary on follow-up studies and "desistance" theories about transgender and gender-nonconforming children

Julia Temple Newhook[a], Jake Pyne[b], Kelley Winters[c], Stephen Feder[d], Cindy Holmes[e], Jemma Tosh[f], Mari-Lynne Sinnott[g], Ally Jamieson[h], and Sarah Pickett[i]

[a]Faculty of Medicine, Janeway Pediatric Research Unit, Memorial University, St. Johns, Newfoundland and Labrador, Canada; [b]School of Social Work, McMaster University, Social Work, Hamilton, Ontario, Canada; [c]GID Reform Advocates, San Diego, California, USA; [d]Department of Pediatrics, University of Ottawa, Ottawa, Canada; [e]School of Social Work, University of Victoria, Victoria, British Columbia, Canada; [f]The Psygentra Institute, White Rock, British Columbia, Canada; [g]Department of Family Medicine, Faculty of Medicine, Memorial University, St. Johns, Newfoundland and Labrador, Canada; [h]Choices for Youth, St. John's, Newfoundland & Labrador, Canada; [i]Faculty of Education, Memorial University, St. Johns, Newfoundland and Labrador, Canada

ABSTRACT

Background: It has been widely suggested that over 80% of transgender children will come to identify as cisgender (i.e., *desist*) as they mature, with the assumption that for this 80%, the trans identity was a temporary "phase." This statistic is used as the scientific rationale for discouraging social transition for pre-pubertal children. This article is a critical commentary on the limitations of this research and a caution against using these studies to develop care recommendations for gender-nonconforming children.

Methods: A critical review methodology is employed to systematically interpret four frequently-cited studies that sought to document identity outcomes for gender-nonconforming children (often referred to as "desistance" research).

Results: Methodological, theoretical, ethical, and interpretive concerns regarding four "desistance" studies are presented. The authors clarify the historical and clinical contexts within which these studies were conducted to deconstruct assumptions in interpretations of the results. The discussion makes distinctions between the specific evidence provided by these studies versus the assumptions that have shaped recommendations for care. The affirmative model is presented as a way to move away from the question of, "How should children's gender identities develop over time?" toward a more useful question: "How should children best be supported as their gender identity develops?"

Conclusion: The tethering of childhood gender diversity to the framework of "desistance" or "persistence" has stifled advancements in our understanding of children's gender in all its complexity. These follow-up studies fall short in helping us understand what children need. As work begins on the 8th version of the *Standards of Care* by the World Professional Association for Transgender Health, we call for a more inclusive conceptual framework that takes children's voices seriously. Listening to children's experiences will enable a more comprehensive understanding of the needs of gender-nonconforming children and provide guidance to scientific and lay communities.

KEYWORDS
Adolescents; children; desistance; dysphoria; follow-up; gender; longitudinal; research; trans; transgender; youth

In the media, among the lay public, and in medical and scientific journals, it has been widely suggested that over 80% of transgender[1] children will come to identify as cisgender[2] once they reach adolescence or early adulthood. This statement largely draws on estimates from four follow-up studies conducted with samples of gender-nonconforming children in one of two clinics in Canada or the Netherlands (Drummond, Bradley, Peterson-Badali, & Zucker, 2008; Steensma, Biemond, de Boer, & Cohen-Kettenis, 2011; Steensma, McGuire, Kreukels, Beekman, & Cohen-Kettenis, 2013; Wallien & Cohen-Kettenis, 2008). This article outlines methodological, theoretical, ethical, and interpretive concerns regarding these studies. We clarify the cultural, historical, and clinical contexts within which these studies were conducted to understand and to deconstruct the embedded foundational assumptions of the research as well as the widespread interpretations of the results. Some of these critiques have been offered elsewhere in

CONTACT Julia Temple Newhook jtemple@mun.ca Janeway Pediatric Research Unit, Faculty of Medicine, Memorial University, 300 Prince Philip Drive, St. John's, NL A1B 3V6, Canada.

commentaries, books, and on-line forums (see Ehrensaft, 2016; Olson & Durwood, 2016; Pyne, 2012, 2014; Serano, 2016; Winters, 2014; Olson, 2016). This analysis seeks to further this discussion by systematically engaging scholarly literature. This team of co-authors encompasses a range of theoretical and disciplinary perspectives which include: clinical care providers in pediatrics, family medicine, and psychology; researchers in the fields of sociology, psychology, neurobiology, and social work; and lived experience as trans and gender diverse people as well as parents and partners of trans and gender diverse people. Our theoretical approach echoes that of the World Professional Association for Transgender Health (Coleman et al., 2012), that transgender identity is "a matter of diversity, not pathology" (p. 4).

We recognize that numerous follow-up studies of gender-nonconforming children have been reported since the mid-20th century (e.g., Green, 1987; Money & Russo, 1979; Zucker & Bradley, 1995; Zuger, 1984). In that era, most research in the domain focused on feminine expression among children assigned male at birth, with the implicit or explicit objective of preventing homosexuality or transsexualism. However, we focus here on the four most recent follow-up studies, published since 2008, which are most often cited as evidence for desistance theories.

Concerns and contributions: What can and cannot be learned from follow-up studies with gender-nonconforming children

Between 2008 and 2013, four follow-up studies of gender-nonconforming children were published in peer-reviewed journals, with samples of children referred to one of two gender clinics in Toronto, Canada or in the Netherlands (Drummond et al., 2008; Steensma et al., 2011, 2013; Wallien & Cohen-Kettenis, 2008). An oft-accepted interpretation of these findings is that approximately 80% of gender-nonconforming children in these studies, by adolescence, identified with their sex assigned at birth. They are often said to have "desisted[3]" " from a prior transgender identity. This presumption links directly to questions about whether to support the self-identities of pre-pubescent gender-nonconforming children. This article reviews these studies and their interpretations by noting a number of positive contributions of this research before systematically exploring the limitations and overgeneralizations that render these studies less reliable than is often assumed.

In Table 1, we summarize the findings and compare several of the main measures and concepts explored in four key peer-reviewed publications between 2008 and 2013, frequently cited to support the 80% "desistance" estimation. These studies included gender-nonconforming children who were referred for clinical care at early ages, and at follow-up in adolescence or early adulthood, categorized them as either "persisters" or "desisters."

In the following commentary, we explore these studies in more depth by focusing on methodological, theoretical, ethical, and interpretive concerns, while noting the extrapolation that may have given rise to problematic treatment recommendations. First, some contributions from the original research are highlighted.

Contributions of follow-up studies with gender-nonconforming children

While this commentary offers critiques of desistance research and its clinical and popular interpretations, these studies have also made contributions to the literature worthy of recognition:

1. qualitative data on trans and gender diverse adolescents in clinical care, and
2. factors in anticipating medical transition for gender-nonconforming children accessing clinical care.

Steensma et al. (2011) offer some unique and valuable qualitative data to this body of literature. In the perspectives of the trans and gender diverse adolescents that are shared in this study, we are able to learn that for these youth, the ages of 10–13 were important for determining which direction their gender would take, and especially significant at this time were their feelings about their bodies and emerging sexualities. The qualitative nature of this research allows us to hear from two youths who had changed their minds about their gender after socially transitioning and who found the process of informing others very stressful. Although this difficulty is not the only possible outcome for youth who make more than one transition (see section on Interpretive Concerns), this research adds to our understanding of the pressures some young people may face when exploring gender in a transphobic society (Steensma et al., 2011).

A second contribution is that the four studies examined offer evidence that statements of transgender

Table 1. Summary of relevant findings across four follow-up studies with gender-nonconforming children (2008–2013).

	Reference	Drummond et al. (2008)	Wallien et al. (2008)	Steensma et al. (2011)	Steensma et al. (2013)
1.	*N*, number of subjects at T0	37 (AFAB)	77	53	127
2.	T0-age	3–12 yrs	5–12 yrs	≤12 yrs	≤12 yrs
3.	Selection by GIDC criteria	DSM-III DSM-III-R DSM-IV	DSM-IV DSM-IV-TR	DSM-IV DSM-IV-TR	DSM-IV DSM-IV-TR
4.	T0-GIDC/*N*	60%	75%	100%	63%
5.	T0-GIDC-subthreshold/*N*	40%	25%	0%	37%
6.	T1-age	≥17 yrs	≥16 yrs	≥14 yrs	≥15 yrs
7.	T1-desistant/*N*	59%	30%	19%	36%
8.	T1-desistant-reported-by-3rd-parties/*N*	n/a	13%	n/a	5%
9.	T1-persistent/*N*	8%	27%	55%	37%
10.	T1-nonbinary	n/a	n/a	2%	n/a
11.	T1-nonparticipant/*N*	32%	30%	25%	22%
12.	Reporting of nonparticipants	Deleted from cohort	Assumed to be desistant	Assumed to be desistant	Assumed to be desistant
13.	Reported desistance rate	88%	73%	45%	63%

Notes: All four studies selected subjects at T0 using the diagnostic criteria for Gender Identity Disorder of Childhood in the DSM-III through the DSM-IV, which did not explicitly require evidence of distress or of gender dysphoria (Row 3). In the Drummond et al. (2008) and Steensma et al. (2013) studies, nearly 40% of subjects were "subthreshold" for GIDC diagnosis, and therefore did not meet those criteria (Row 4). In Row 9, rates are shown for persistent GIDC diagnosis or distress of gender dysphoria confirmed by re-assessment at T1 in adolescence or young adulthood. Far fewer subjects were confirmed with desisted GIDC diagnosis or with desisted gender dysphoria distress by re-assessment at T1 (Row 7), than were reported as desisters in these papers (Row 13). Five percent of Steensma et al. (2013) subjects were judged desistant based only on parental or other 3rd party information at T1, with no reported examination of possible bias among those parties (Row 8). One subject in Steensma et al. (2011) identified as nonbinary at T1 but was categorized as desistant (Row 10). Between 1/5th and 1/3rd of the samples did not participate in follow-up evaluation at T1 (Row 11). In Steensma (2013), 22% of subjects were termed "nonresponders" at T1, including 19% whose status was altogether unknown and 3% who were reported as "indicated" desistant, though not confirmed by actual participation at T1 re-assessment (p. 584). Despite their outcome status being unknown or unconfirmed, they were categorized as desisters in all three Dutch studies, Steensma et al. (2011) and Steensma et al. (2013), Wallien and Cohen-Kettenis (2008) (Row 12). Desistance rates in the four papers were reported by the authors as follows: T1-desistant/T1-participant for Drummond et al. (2008), where T1 nonparticipants were deleted from the cohort denominator; (T1-desistant + T1-nonparticipant)/*N* for Wallien and Cohen-Kettenis (2008); (T2-desistant + T2-nonbinary + T2-nonparticipant)/*N* for Steensma et al. (2011); and (T2-desistant + T2-desistant-reported-by-3rd-parties + T2-nonparticipant)/*N* for Steensma et al. (2013).

identity in childhood may help to anticipate transgender identity in adolescence or adulthood, and similarly, that the reported intensity of feelings of gender dysphoria in childhood may help to anticipate feelings of gender dysphoria in adulthood. Drummond et al. (2008) found that the two participants classified as gender dysphoric at follow-up "recalled significantly more cross-gender identity and role behavior in childhood than participants classified as having no gender dysphoria" (p. 41). Wallien and Cohen-Kettenis (2008) reported that "all participants in the persistence group were given a complete GID diagnosis in childhood, whereas half of the group of desisting children was subthreshold for the diagnosis" (pp. 1420–1421). Finally, Steensma et al. (2013) reported that "explicitly asking children with GD with which sex they identify seems to be of great value in predicting a future outcome for [children diagnosed with gender dysphoria]." (p. 588). Both media and scientific discussion of this research have tended to downplay what is suggested here—the value of asking a child about their gender identity. Gender identity can indeed shift and evolve over time, and thus a young person who did not express trans identity in childhood should not be dismissed in their teen years on this basis, yet the persistence of this stated identity for some youth may be instructive. Regardless of predictive value, however, current approaches to care recommend that care providers prioritize young people's stated identities, perceptions, and needs in the present moment, as opposed to attempting to estimate the likelihood of future identity and needs (Hidalgo et al., 2013; Temple-Newhook et al., in press 2018). As clinicians Schreier and Ehrensaft (2016) suggest: "Want to know a child's gender? Ask." We now turn to outlining a series of critiques and concerns regarding these studies, including: methodological, theoretical, ethical, and interpretive concerns.

Methodological concerns

We have identified the following methodological concerns in these four studies:

1. the potential misclassification of child research participants
2. the lack of acknowledgement of social context for research participants
3. the age of participants at follow-up, and
4. the potential misclassification of adolescent and young adult participants lost to follow-up.

The first two methodological concerns address the broad inclusion criteria for those studied in childhood. Rather than a representative group of transgender

children, which is assumed in many interpretations, this literature focused on small groups of gender-nonconforming children in two clinics. In Table 1, some but not all of these children were diagnosed in childhood with gender identity disorder in children (GIDC), in prior editions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) published by the American Psychiatric Association (1980, 1987, 1994, 2000). The GIDC category was replaced by gender dysphoria in children (GDC) in the DSM-5 (American Psychiatric Association, 2013). Gender dysphoria was not defined as a diagnostic category until release of the DSM-5 in 2013, and therefore, no subjects were formally diagnosed with gender dysphoria in childhood or at follow-up in any of these studies. However, the plain-language meaning of gender dysphoria, as distress regarding incongruent physical sex characteristics or ascribed social gender roles, has been established since the 1970s (Fisk, 1973). When these four studies refer to gender dysphoria, they are referring to this plain-language context of distress, and not the newer DSM-5 diagnostic category.

Due to such shifting diagnostic categories and inclusion criteria over time, these studies included children who, by current DSM-5 standards, would not likely have been categorized as transgender (i.e., they would not meet the criteria for gender dysphoria) and therefore, it is not surprising that they would not identify as transgender at follow-up. Current criteria require identification with a gender other than what was assigned at birth, which was not a necessity in prior versions of the diagnosis. For example, in Drummond et al. (2008) study (Table 1), the sample consisted of many children diagnosed with GIDC, as defined in the DSM editions III, III-R, and IV (American Psychiatric Association, 1980, 1987, 1994). Yet the early GIDC category included a broad range of gender-nonconforming behaviors that children might display for a variety of reasons, and not necessarily because they identified as another gender. Evidence of the actual distress of gender dysphoria, defined as distress with physical sex characteristics or associated social gender roles (Fisk, 1973), was dropped as a requirement for GIDC diagnosis in the DSM-IV (American Psychiatric Association, 1994; Bradley et al., 1991). Moreover, it is often overlooked that 40% of the child participants did not even meet the then-current DSM-IV diagnostic criteria. The authors conceded: "…it is conceivable that the childhood criteria for GID may 'scoop in' girls who are at relatively low risk for adolescent/adult gender-dysphoria" and that "40% of the girls were not judged to have met the complete DSM criteria for GID at the time of childhood assessment… it could be argued that if some of the girls were subthreshold for GID in childhood, then one might assume that they would not be at risk for GID in adolescence or adulthood" (p. 42). By not distinguishing between gender-nonconforming and transgender subjects, there emerges a significant risk of inflation when reporting that a large proportion of "transgender" children had desisted. As noted by Ehrensaft (2016) and Winters (2014), those young people who did not show indications of identifying as transgender as children would consequently not be expected to identify as transgender later, and hence in much public use of this data there has been a troubling overestimation of desistance.

The second concern brings attention to the scope of the four studies discussed. Inferences from clinical research are always bound to specific locations and timeframes. Generalizing from research on gender identity is particularly problematic because notions of gender are highly dependent on social and historic context. For example, the meaning of gender-conformity and nonconformity varies greatly, and in some non-Western and indigenous cultures, gender diversity is either celebrated or considered nonproblematic (Driskell, Finley, Gilley, & Morgensen, 2011; Hunt, 2016). Furthermore, even within Toronto and the Netherlands, this research was limited to children whose parents chose to bring them to a clinic for diagnosis and treatment and thus may have believed the child's difference was a problem, and one that required psychological treatment. Children whose parents affirmed their gender (or who did not wish to or who were unable to access clinical treatment for any reason) were likely not included in these studies. This is significant because more recent work has shown that children raised by parents who validate their gender identity (Durwood, McLaughlin, & Olson, 2017; Olson & Durwood, 2016) are likely to demonstrate a different (and in some respects healthier) life course than children with parents who are reluctant or unwilling to affirm gender-nonconformity. The outcomes in the one group, therefore, may not generalize to the other.

The third methodological concern centers on the age at which follow-up was conducted. As noted in Table 1, only a minority of the young people who consented to

be re-studied were diagnosed in adolescence with gender identity disorder in adolescents or adults (GIDAA) and/or chose to undergo certain trans-affirming surgeries in early adulthood. Yet in these four studies (Table 1), the mean age at follow-up ranged from 16.04 (Steensma et al., 2013) to 23.2 years (Drummond et al., 2008) and included adolescents as young as 14 years (Steensma et al., 2011). It is important to acknowledge that this represents a very early follow-up point in an individual's life, and that a trans person might assert or reassert their identity at any point in their life. An assumption has been made that young people not diagnosed with GID (or Gender Dysphoria in the current DSM-5) by late adolescence and/or not pursuing medical transition by a relatively early age, can then by default be "correctly" categorized as cisgender for their lifetime. However, this conclusion is contradicted when an unknown number of those counted as "desisters" may transition later, after the point of follow-up. Research has found that many trans-identified individuals come out or transition later in adulthood (Reed, Rhodes, Schofield, & Wylie, 2009).

A fourth methodological concern focuses on the misclassification of participants who did not participate in follow-up. A significant challenge to any longitudinal study is that a number of original participants will not be able to be located or will not provide research consent even if located, and so will be lost to follow-up (attrition). One can only continue to study those participants who can be located and contacted, and who also then consent to re-enrol. High levels of attrition limit the generalizability of longitudinal research findings, particularly when attrition might be non-random and related to the outcome. In the Drummond et al. (2008) study, the data from participants who were lost to follow-up (32% of participants) were removed from the study and analyses were conducted only with participants who returned. This *list deletion* method of handling attrition can be risky and can introduce bias, as Deng, Hillygus, Reiter, Si, and Zheng (2013) note, because there is much scholarship suggesting that certain individuals are more likely to be lost to follow-up than others.

The three other studies analyzed in this review (Steensma et al., 2011, 2013; Wallien & Cohen-Kettenis, 2008) took what might be understood as an even riskier approach to handling attrition, by imputing all outcomes to missing participants as "desisters," in effect, venturing to guess about data that was unavailable. In these studies, 30% of Wallien and Cohen-Kettenis' (2008), 25% of Steensma, et al.'s (2011), and 22% of Steensma et al.'s (2013) participants who did not respond or did not participate at adolescent follow-up were counted as desisters (Table 1). In explanation, Steensma et al. (2011) write: "As the Amsterdam Gender Identity Clinic for children and adolescents is the only one in the country, we assumed that their gender dysphoric feelings had desisted, and that they no longer had a desire for sex reassignment." (p. 501). In other words, desistance was assessed based on whether or not participants re-engaged with this specific clinic by a specific time. This methodological choice neglects a number of important considerations: (1) the fact that not all transgender people wish to medically transition, yet still identify as trans; (2) the socio-economic or cultural factors that may influence whether an adolescent seeks psychological or medical treatment; (3) the possibility of a negative perception of the initial clinic experience, which might discourage a youth's return; (4) the possibility of a youth moving out of the country, being institutionalized in a mental health facility or even the possibility of death (including suicide), none of which negate a trans identity; and, (5) the possibility that some young people might repress their gender identity for a period of time, due to societal transphobia, family rejection, safety, employment and housing security, or pressure from therapies designed to discourage trans identity (Kennedy & Hellen, 2010). The phenomenon of realizing one's gender identity long before expressing it to others has been illustrated in the Trans PULSE study conducted in Ontario, Canada. While 59% of participants had socially transitioned within the four years prior to study, the majority of participants first realized that they were trans before the age of 10 years (Scheim & Bauer, 2015). The classification (and potential misclassification) of participants lost to follow-up as desisters could result in a significant overestimation of the number of young adults assumed to be cisgender.

Theoretical concerns

We have identified the following theoretical concerns in the four studies:

1. assumptions inherent in "desistance" terminology
2. binary gender framework, and
3. presumption of gender stability as a positive outcome.

The first theoretical concern pertains to the unnecessary conceptualizing of shifts in gender identity as either "persistence" or "desistance." The etymology of

the word desistance (from the Latin desistere, meaning to stop or cease) reveals that the dominant framework for understanding variations of childhood gender is rooted in the field of criminology (e.g., Farrall, Bottoms, & Shapland, 2010; Sampson & Laub, 2003; Stouthamer-Loeber, Wei, Lober, & Masten, 2004), where desistance is defined as "the cessation of offending or other antisocial behavior" (Kazemian, 2011, p. 656). This choice of terminology positions gender identity development as a pathway of either "normal" or "deviant" identity. In addition, the use of the term desistance in all four studies positions cisgender and transgender as immutable discrete categories (Serano, 2016). In this framework, cisgender identity tends to be seen as the healthy opposite of a problematic transgender identity. Assertion of a cisgender identity at any point in the life cycle is often assumed to be valid and invalidates any previous assertion of transgender identity; yet a transgender identity is only viewed as valid if it is static and unwavering throughout the life course and if it emerges in a particular time period (the period of study). In our research and practice experience, a rigid categorization of gender does not reflect the lived experiences of transgender and gender-nonconforming children. A child who has identified as transgender may indeed at some point in their life assert their birth-assigned gender, but this is not necessarily the end of their gender journey.

A second theoretical concern is that the terminology of "desistance" depends on a binary understanding of gender. Each of the four studies used binary language to refer to children as "boys and girls," prioritizing the sex they were assigned at birth, as opposed to their own identity. Furthermore, this language makes nonbinary and intersex identities invisible. For example, Steensma et al. (2011) define the following individual as a desister and thus cisgender, in spite of the young person's own self-identification: "At the time of the interview, Desister 1 (18 years of age), still desired to be a woman, with breasts and the possibility of giving birth. However, he considered himself 50% male and 50% female" (p. 512). The authors acknowledge that not every person who experiences gender dysphoria will seek medical transition or assert a binary gender identity. They add that "[i]t would be worthwhile to follow [this nonbinary-identifying young person's] development much longer, to see whether [their] ambiguous gender feelings were just part of a passing phase (either into desistence or persistence) or whether they remained a stable characteristic of this person" (p. 513). We question here the characterization of a self-described "50% male and 50% female" research participant as "ambiguous," instead of a term supplied by the participant, and hope that if this study were conducted today, there would be greater recognition of nonbinary gender identities.

A third theoretical concern is the embedded assumption in these studies that "stability" of gender identity is a positive health outcome that should be prioritized for all children. A desistance framework reinforces a static understanding of gender that hinders us from understanding the experience of a child whose gender identity is more fluid, or changeable over time. While the current understanding of the developmental trajectories of gender state that *most* children are aware of their gender identity by the age of 4 years (American Association of Pediatrics, Human Rights Commission, & American College of Osteopathic Pediatricians, 2016), this does not suggest that those children for whom gender identity is more fluid or slower to develop are not also following a healthy developmental trajectory. While many individuals experience their gender identity as stable throughout their lifetimes, others find that a gender that "fits" at age four may be different from what fits at age seven, age 18, or age 65. None of these identities are "wrong"; instead they may have been perfectly and precisely the right fit for that person at that moment. Further, for some individuals the most consistent aspect of their gender is that it is fluid or ever-changing. Many individuals move through a process of exploration and/or "[renegotiation] of one's gender throughout childhood or adulthood with no observable detriment to their mental health" (Ehrensaft, 2016, p. 59). An alternative framework would conceptualize changes and developments in gender identity not as errors in the development of a "true" gender, but as necessary paths of exploration along a journey of self-discovery that might be lifelong. Finally, as Bryant (cited in Schwartzapfel, 2013) points out, it is likely that the future identities of today's children cannot be known for certain, given that the language to describe or acknowledge these identities may not yet exist. There is no evidence that caring for a child in the present requires knowing their future adult gender identity. A longitudinal research design that records identity at two relatively early intervals is therefore arguably not the most appropriate tool for understanding either children's or adult's health needs.

The dominance of binary language in interpretations of desistance research is implied in statements such as "the majority of those who desist by or during adolescence grow up to be gay, not transgender" (Drescher & Pula, 2014, p. S18). Such statements conflate gender identity and sexual orientation, with an underlying assumption that the options are between identifying as gay and cisgender, or as transgender. Framing research on childhood gender diversity in terms of desistance and persistence tends to reproduce and reinforce this limited binary perspective on gender and sexuality. The conception of gender reflected in these studies represents a historical and cultural moment that differs from both traditional *and* current understandings of gender and sexuality. As scholars of human sciences, we are reminded that clinical and research disciplines are human-made frameworks to understand complex identities and actions. In scientific and medical research, it is important to acknowledge that the categories used to study people (e.g., desisters) often reflect the assumptions and beliefs of the researchers themselves. We suggest that if we find that people do not fit our categories, then it is the categories that must change.

Ethical concerns

We have also identified ethical concerns in these four studies:

1. intensive treatment and testing of child participants,
2. questionable goals of treatment, and
3. lack of consideration of children's autonomy.

From an ethical perspective, it is important to consider that research itself is an intervention. These studies took place in the context of gender clinics in which children were put through a substantial degree of testing over periods of months or years. For example, Drummond et al. (2008) report that in their study, children and their parents were administered: the Draw a Person test; a free-play task; the Playmate and Playstyle Preferences Structured Interview; sex-typed responses on the Rorschach test; the Gender Identity Questionnaire for Children; a measure of activity level/extraversion; and the Games Inventory. Critiques of the practice of diagnosing gender-nonconforming children (with the GIDC diagnosis from DSM-III, DSM III-R, DSM IV) began to be published in the late 1990s and argued that healthy children might have their self-esteem damaged and their trust in therapy eroded by being brought into stigmatizing diagnostic and treatment settings (Isay, 1997; Langer & Martin, 2004; Menvielle, 1998; Pickstone-Taylor, 2003; Vanderburgh, 2009). This concern continues and highlights the need for research into possible adverse effects and ethical complications related to extensive and ongoing psychological testing for children in clinical settings.

A second ethical concern is that many of the children in the Toronto studies (Drummond et al., 2008; Zucker & Bradley, 1995) were enrolled in a treatment program that sought to "lower the odds" that they would grow up to be transgender (Drescher & Pula, 2014; Zucker, Wood, Singh, & Bradley, 2012; Paterson, 2015). Zucker et al. (2012) wrote: "…in our clinic, treatment is recommended to reduce the likelihood of GID persistence" (p. 393). In a Hastings Centre Report on LGBT Bioethics, Drescher and Pula (2014) explain the Toronto clinic's approach: "The clinic claims its approach decreases the likelihood that GD will persist into adolescence, leading to adult transsexualism, which, for various reasons, such as social stigma and a lifetime of medical treatment, is an outcome the clinic considers undesirable" (pp. S17–18). Drescher and Pula (2014) make an ethical inquiry about this approach: "Since no clinician can accurately predict the future gender identity of any particular child, shouldn't we assume that efforts to discourage cross-gender play and identifications may be experienced as hurtful and possibly even traumatic, since, for some children, gender dysphoria will persist into adolescence and adulthood?" (p. S19). Drescher and Pula elaborate: "Are the harms so unknown or so great that it is unethical to offer such treatment at all?" (p. S19). The Toronto clinic was closed in 2015 (Schreier & Ehrensaft, 2016), but questions regarding the interpretation of research conducted in this setting are ongoing.

Drummond et al. (2008) report that their follow-up study provides information on the "natural histories" of "girls with gender identity disorder" (p. 34), yet the clinical pursuit of an a priori goal for a child's gender is already inconsistent with the meaning of the term "natural history," which refers to the natural progression of a condition in the absence of treatment (Center for Disease Control, 2012). While the Netherlands clinic did not discourage children from exploring their gender expression, it did discourage children from

socially transitioning prior to puberty (Drescher & Pula, 2014; Steensma et al., 2013). It is important to acknowledge that discouraging social transition is itself an intervention with the potential to impact research findings, as discussed below in reference to interpretive concerns.

A lack of consideration of children's autonomy in desistance literature is a third ethical concern. Children have their own rights to autonomy and self-determination (Powell, Fitzgerald, Taylor, & Graham, 2012).[4] However, children's own assertions of identity and their own perspectives on their gender are subordinate in this literature to the diagnostic measures created by clinicians and researchers. In one case, Steensma et al. (2013) write that, "because the role of parental report on gender-variant behaviors and surface behaviors such as gender role transitioning are of less value in predicting a future persistence of gender dysphoria in [children assigned female at birth], it seems important to provide extra focus on [assigned-female children's] own experiences of cross-gender identification and wishes" (p. 588). We agree with this statement of concern for children's own identification and wishes and would extend it to children of all genders.

This consideration of children's own wishes should also extend to their right to decline participation in research. In the four studies, there is an absence of information about whether research participation was optional and if steps were taken to ensure that children could decline research consent while continuing to receive needed services. The need to decouple research participation from access to medical and mental health care is consistent with emerging proposals for ethical research with transgender subjects (Adams et al., 2017; Devor, Bauer, Pyne, Heinz, & Marshall, 2016)

These ethical concerns raise questions about the validity of research with children whose parents believe they have a medical problem, who are subjected to a high level of testing and treatment, who are disallowed or discouraged from asserting their own gender identity, and who are being raised in a broader society that often punishes perceived transgressions of male and female boundaries. Interpretations of desistance research have assumed that the difference recorded between measures of gender dysphoria and/or identity at childhood and at early adulthood mean that "the majority of [trans children] will become comfortable with their natal gender over time" (Byne et al., 2012). However, the larger social context shaping young people's identities is essential to consider. Wallien and Cohen-Kettenis (2008) point out that there are challenges in research based on self-report of sexual orientation, given that "social desirability is a key validity issue in the assessment of sexual orientation during the adolescent years" (p. 1421). It would be fair to assume that the same concern would hold for self-report of gender identity. Drummond et al. (2008) attempt to account for the possible effect of social desirability by assessing participant responses to questions about a range of socially undesirable issues, yet it is unclear if this can account for the way transgender identity might be uniquely undesirable in a clinic that explicitly seeks to discourage it (see Zucker et al., 2012). In interpreting the results of these studies, it is important to ask questions about limitations in the validity of self-report when the research is conducted under conditions that might compromise authentic responses, for example, within a clinic where transgender identity is defined as less desirable than cisgender identity.

Interpretive concerns

We also have concerns with the authors' interpretation in these four studies, including:

1. the assumption that unknown future adult needs should supersede known childhood needs, and
2. the underestimation of harm when attempting to delay or defer transition.

Desistance studies are often drawn on to suggest that delaying a young person's social transition is justified because it may prevent them from having to transition back in the future. There is an assumption that a second transition would be distressing. Steensma et al. (2013) write: "the percentage of transitioned children is increasing …which could result in a larger proportion of children who have to change back to their original gender role, because of desisting GD, accompanied with a possible struggle" (pp. 588–589). Yet we note that this projected struggle is acknowledged only as "possible" rather than certain (p. 589). Similarly, in a letter to the editor entitled "Gender Transitioning before Puberty?", Steensma and Cohen-Kettenis (2011) write: "It is conceivable that the drawbacks of having to wait until early

adolescence (but with support in coping with the gender variance until that phase) may be less serious than having to make a social transition twice" (p. 649). Yet again, this statement itself acknowledges that future distress is merely "conceivable" and again, not certain. As Ehrensaft, Giammattei, Storck, Tishelman, and Keo-Meier (2018) note, the evidence that a second transition would be traumatic is very thin, drawn from a case study of two children who found a reversion back to their original gender challenging in Steensma and Cohen-Kettenis's (2011) clinic. Yet in another clinic (Edwards-Leeper & Spack, 2012), a de-transitioning girl and her mother expressed gratitude for her opportunity to live as a boy for a time, and they felt that if she had been forced to live as a girl for her entire childhood, that her mental health would have suffered. Thus, with many possible outcomes for the future, young people's needs in the present must be prioritized.

A further related interpretive concern is the presumption that childhood needs, adolescent needs, and adult needs should "match" in a simplistic sense. Yet a child may need to use "she" pronouns in childhood, "he" pronouns in adolescence, and "they" pronouns in adulthood. Nothing about this is inherently problematic. What is needed in childhood may differ from what is needed in adolescence or adulthood, but this does not negate childhood needs. What is problematic is the assumption that a potential future shift in a child's gender identity is a justification for suppressing or redirecting their assertion of identity in childhood.

The underestimation of harm in suppressing or redirecting children's gender expression is the most serious concern in interpretations of desistance literature. That gender identity or expression may change among children (or adults) does not support the hypothesis that it is preferable or possible to externally "coax" gender in a particular direction, or that this could be done without harm. In contrast, the positioning of this goal as benign ignores the potential harms to young people who have undergone such treatments (Bryant, 2006). A 2013 attachment-based theoretical comparison of gender therapies for children concluded that there is a risk that children who are discouraged from expressing their gender identity may integrate shame into their fundamental sense of self (Wallace & Russell, 2013). Drescher and Pula (2014) offer: "It could be construed … that clinical attempts to prevent transsexualism, no matter how well meaning, are unethical because they demean the dignity of gender-variant children" (p. S19). Although the term "conversion therapy" originally referred to religious-based therapies purported to change an individual's sexual orientation, as noted by the American Academy of Child and Adolescent Psychiatry (AACAP), the meaning of the term has expanded in recent years to encompass efforts to change an individual's core gender identity or promote a preferred outcome for their gender identity, therapies that according to the AACAP "lack scientific credibility" (AACAP, 2018). In our experience, disallowing children's assertions of gender identity is far from a "neutral" option. From a developmental perspective, a child who is repeatedly discouraged when she earnestly insists on being called "she," is learning, on a fundamental level, that (1) she cannot trust her own knowledge of herself and, (2) the adults she depends on may not value her for who she knows herself to be.

Lastly, while many clinicians would not propose attempting to alter gender expression, many still interpret desistance research as support for delaying transition, lest a trans identity becomes more likely. Steensma et al. (2013) write: "…with a link between social transitioning and the cognitive representation of the self [social transition may] influence the future rates of persistence" (pp. 588–589). Yet we would ask why an increase in the number of transgender people ("persistence") would be interpreted in a negative light, and how this sentiment could be consistent with the WPATH position that transgender identity is a matter of diversity not pathology (Coleman et al., 2012). Drescher and Pula (2014) as well as Ehrensaft et al. (2018) note that at times there appears to be a willingness to expose transgender children to the stress of living in a gender they do not identify with, in order to protect cisgender children from the possibility of "mistakenly" transitioning. Yet we would contend that the quality of life of transgender children is no less important and no less valuable than that of cisgender children.

Discussion: What is the future of care for trans and gender diverse children?

It is essential to distinguish between the evidence provided by these studies and the flawed interpretations of these studies that may be used to shape care for trans and gender-nonconforming children. These concerns address the differences between the questions: (1) *how should children's gender identities develop over*

time? and, (2) *how should children best be supported as their gender identity develops?*

In the first question, desistance studies give us some information about how some gender-nonconforming children's identities have been recorded in certain cultural and clinical circumstances. However, the World Professional Association for Transgender Health *Standards of Care* now recognizes gender-nonconformity as a matter of diversity not pathology (Coleman et al., 2012, pp. 1, 3, 4, 6); thus, no path need be considered the correct or healthy trajectory. While transgender identity has certainly been pathologized in Western medicine, present-day best practice promotes the recognition of a plurality of healthy developmental trajectories for gender.

In the second question, the affirmative care model, which is now practiced by the majority of North American gender clinics (Ehrensaft, 2016), promotes support for children as their gender identity develops, with no expectations for any particular direction of the gender journey. In the affirmative care model, children are provided with the space to explore and try out different self-expressions to discover a place that is comfortable for them (Ehrensaft, 2016; Hidalgo et al., 2013). This means that instead of attempting to direct a child toward a particular identity, parents and caregivers accept a child's own individual journey. Within this model of care, adult scrutiny and investment in any particular current or future gender identity are removed. Children are not prevented from exploring aspects of gender as they develop a sense of what fits for them through the language available to them at that time. Within this model of care, it is understood that the gender that is the "right fit" may differ at different ages and stages of life. Emergent research on the health and well-being of trans children who are affirmed in their gender identity, indicates mental health outcomes equivalent with cisgender peers (Durwood et al., 2017; Olson, Durwood, DeMeules, & McLaughlin, 2016). As Sherer (2016) and Turban (2017) note, this is in stark contrast to the high levels of psychological distress and behavioral problems documented among children who were discouraged from asserting their identities in childhood (Cohen-Kettenis, Owen, Kaijser, Bradley, & Zucker, 2003).

From a research perspective, noting the gap between the questions of (1) gender trajectories and (2) the health of gender-nonconforming children, we suggest that longitudinal studies about identity "desistance" or "persistence" are not the best tools for understanding the needs of gender-nonconforming children. As work begins on the 8th version of the *Standards of Care* by the World Professional Association for Transgender Health, we call for a refocus of research and clinical practice with transgender and gender-nonconforming children and youth, to prioritize listening to how young people articulate their wishes and needs. The potential harm inherent in approaches that lack support for children's own assertions must be acknowledged as we create a more nuanced framework for understanding and caring for gender-nonconforming children. Such a framework would integrate an understanding of intersectionality (Cole, 2009; Collins, 1990; Crenshaw, 1991; Singh, 2013), which involves taking into account the multiple systems of oppression that simultaneously shape trans and gender-nonconforming children's lives.

Conclusion

In this critical review of four primary follow-up studies with gender-nonconforming children in Toronto, Canada and the Netherlands (Table 1), we identify a total of 12 methodological, theoretical, ethical, and interpretive concerns as well as two often-overlooked contributions of this literature. We conclude that, while our understanding of gender diversity in adults has progressed, the tethering of childhood gender identity to the idea of "desistance" has stifled similar advancements in our understanding of children's gender diversity. As we progress towards a fuller understanding of children's gender in all its complexity, it will be important to move beyond longitudinal studies of identity that seek to predict children's futures, and instead prioritize respect for children's autonomy in the present. For all the resources devoted to studying these children, we have much more to learn by listening to them.

Notes

1. We use the term trans or transgender when an individual's gender differs from the one assigned to them at birth and/or differs from what others expect of their physical presentation. This is not a universal definition, rather it is a description of our use of the term in this article.
2. We use the term cis or cisgender when an individual's gender aligns with the one assigned to them at birth and matches what others expect of their physical presentation.

3. We consider the term "desistance" to be flawed (see *Theoretical Concerns* section) but we use the term in this commentary because it is widely understood in the field of transgender health.
4. The authors acknowledge that beliefs about children's autonomy vary cross-culturally.

Declaration of conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Acknowledgments

The authors would like to thank Ayden Scheim and Greta Bauer as well as the anonymous reviewers and editors of this special issue for their helpful comments on earlier drafts.

Funding

No funding was provided for the writing of this article.

References

Adams, N., Pearce, R., Veale, J., Radix, A., Castro, D., Sarkar, A., & Thom, K. C. (2017). Guidance and ethical considerations for undertaking transgender health research and institutional review boards adjudicating this research. *Transgender Health*, *2*(1), 165–174. doi:10.1089/trgh.2017.0012

American Academy of Child & Adolescent Psychiatry. (2018, February). *Conversion Therapy*. Retrieved from https://www.aacap.org/AACAP/Policy_Statements/2018/Conversion_Therapy.aspx

American Association of Pediatrics, Human Rights Commission, & American College of Osteopathic Pediatricians. (2016). *Supporting and caring for transgender children*. Washington, DC: Human Rights Commission. Retrieved from http://www.hrc.org/resources/supporting-caring-for-transgender-children

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. rev). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed. rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed. rev.). Washington, DC: Author.

Bradley, S., Blanchard, R., Coates, S., Green, R., Levine, S., Meyer-Bahlburg, H., … Zucker, K. (1991). Interim report of the DSM-IV subcommittee on gender identity disorders. *Archives of Sexual Behavior*, *20*, 333–343. doi:10.1007/BF01542614

Bryant, K. (2006). Making gender identity disorder of childhood: Historical lessons for contemporary debates. *Sexuality Research and Social Policy*, *3*, 23–39. Retrieved from https://doi.org/10.1525/srsp.2006.3.3.23. doi:10.1525/srsp.2006.3.3.23

Byne, W., Bradley, S., Coleman, E., Eyler, A. E., Green, R., Menvielle, E. J., … Tompkins, D. A. (2012). Report of the American psychiatric association task force on treatment of gender identity disorder. *Archives of Sexual Behavior*, *41*, 759–796. doi:10.1007/s10508-012-9975-x

Center for Disease Control. (2012). *Principles of epidemiology in public health practice*. Retrieved from https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section9.html

Cohen-Kettenis, P. T., Owen, A., Kaijser, V. G., Bradley, S. J., & Zucker, K. J. (2003). Demographic characteristics, social competence, and behavior problems in children with gender identity disorder: A cross-national, cross-clinic comparative analysis. *Journal of Abnormal Child Psychology*, *31*, 41–53. doi:10.1023/A:1021769215342

Cole, E. (2009). Intersectionality and research in psychology. *American Psychologist*, *64*, 170–180. doi:10.1037/a0014564

Coleman, E., Bockting, W., Botzer, M., Cohen-Kettenis, P., DeCuypere, G., Feldman, J., … Zucker, K. (2012). Standards of care for the health of transsexual, transgender, and gender- nonconforming people (Ver. 7). *International Journal of Transgenderism*, *13*(4), 165–232. doi:10.1080/15532739.2011.700873

Collins, P. (1990). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. Boston: Unwin Hyman.

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, *43*, 1241–1299. doi:10.2307/1229039

Deng, Y., Hillygus, D., Reiter, J., Si, Y., & Zheng, S. (2013). Handling attrition in longitudinal studies: The case for refreshment samples. *Statistical Science*, *28*, 238–256. doi:10.1214/13-STS414

Devor, A., Bauer, G., Pyne, J., Heinz, M., & Marshall, Z. (2016, June). *Generating national guidelines for research involving transgender people*. Paper presented at the World Professional Association for Transgender Health Biennial Symposium, Amsterdam, ND.

Driskell, Q., Finley, C., Gilley, B. J., & Morgensen, S. L. (2011). *Queer indigenous studies: Critical interventions in theory, politics and literature*. Tuscon, AR: University of Arizona Press.

Drescher, J., & Pula, J. (2014). Ethical issues raised by the treatment of gender-variant prepubescent children. *LGBT Bioethics: Visibility, Disparities, and Dialogue, special report, Hastings Center Report*, *44*, S17–S22.

Drummond, K., Bradley, S., Peterson-Badali, M., & Zucker, K. (2008). A follow-up study of girls with gender identity disorder. *Developmental Psychology*, *44*, 34–45. doi:10.1037/0012-1649.44.1.34

Durwood, L., McLaughlin, K., & Olson, K. (2017). Mental health and self-worth in socially transitioned transgender children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(2), 116–123. doi:10.1016/j.jaac.2016.10.016

Edwards-Leeper, L., & Spack, N. P. (2012). Psychological evaluation and medical treatment of transgender youth in an interdisciplinary "Gender Management Service" (GeMS) in a major pediatric center. *Journal of Homosexuality*, *59*, 321–336. doi:10.1080/00918369.2012.653302

Ehrensaft, D. (2016). *The gender creative child*. New York, NY: The Experiment Publishing.

Ehrensaft, D., Giammattei, S., Storck, K., Tishelman, A., & Keo-Meier, C. (2018). Prepubertal social gender transitions: What we know; what we can learn—A view from a gender affirmative lens. *International Journal of Transgenderism*. doi:10.1080/15532739.2017.1414649.

Farrall, S., Bottoms, A., & Shapland, J. (2010). Social structures and desistance from crime. *European Journal of Criminology*, *7*, 546–570. doi:10.1177/1477370810376574

Fisk, N. (1973). Gender dysphoria syndrome. (The how, what, and why of a disease). In D. Laub & P. Gandy (Eds.), *Proceedings of the second interdisciplinary symposium on gender dysphoria syndrome* (pp. 7–14). Palo Alto, CA: Stanford University Press.

Green, R. (1987). *The "sissy boy syndrome" and the development of homosexuality*. New Haven, CT: Yale University Press.

Hidalgo, M., Ehrensaft, D., Tishelman, A., Clark, L., Garofalo, R., Rosenthal, S., … Olson, J. (2013). The gender affirmative model: What we know and what we aim to learn. *Human Development*, *56*, 285–290. doi:10.1159/000355235

Hunt, S. (2016). *An introduction to the health of two-spirit people: Historical, contemporary and emergent issues*. Prince George, BC: National Collaborating Centre for Aboriginal Health. Retrieved from https://www.ccnsa-nccah.ca/docs/emerging/RPT-HealthTwoSpirit-Hunt-EN.pdf

Isay, R. A. (1997). Remove gender identity disorder from DSM. *Psychiatric News*, *32*, p. 9. Retrieved from http://psychnews.org/pnews/97-11-21/isay.html

Kazemian, L. (2011). Desistance from crime and delinquency. In R. Levesque (ed.) *Encyclopedia of Adolescence* (pp. 656–664). New York: Springer-Verlag.

Kennedy, N., & Hellen, M. (2010). Transgender children: more than a theoretical challenge. *Graduate Journal of Social Science*, *7*(2), 25–43.

Langer, S. J., & Martin, J. I. (2004). How dresses can make you mentally ill: Examining gender identity disorder in children. *Child and Adolescent Social Work Journal*, *21*, 5–23. doi:10.1023/B:CASW.0000012346.80025.f7

Menvielle, E. (1998). Gender identity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*(3), 243–244. doi:10.1097/00004583-199803000-00001

Money, J., & Russo, A. J. (1979). Homosexual outcome of discordant gender identity/role: Longitudinal follow-up. *Journal of Pediatric Psychology*, *4*, 29–41. doi:10.1093/jpepsy/4.1.29

Olson, K. R. (2016). Prepubescent transgender children: What we do and do not know. *Journal of the American Academy of Child and Adolescent Psychiatry*, *55*, 155–156. doi:10.1016/j.jaac.2015.11.015

Olson, K., & Durwood, L. (2016, January 14). Are parents rushing to turn their boys into girls? *Slate*. Retrieved from http://www.slate.com/blogs/outward/2016/01/14/what_alarmist_articles_about_transgender_children_get_wrong.html

Olson, K., Durwood, L., DeMeules, M., & McLaughlin, K. (2016). Mental health of transgender children who are supported in their identities. *Pediatrics*, *137*(3), 1–8. doi:10.1542/peds.2015-3223

Paterson, T. (2015, February 21). As trans issues become mainstream, question of how to address variant gender expression comes to forefront. National Post. Retrieved from http://news.nationalpost.com/life/as-trans-issues-become-mainstream-question-of-how-to-address-variant-gender-expression-comes-to-forefront

Pickstone-Taylor, S. (2003). Children with gender nonconformity. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*, 266. doi:10.1097/00004583-200303000-00005

Powell, M. A., Fitzgerald, R., Taylor, N. J., & Graham, A. (2012). International literature review: Ethical issues in undertaking research with children and young people. *Literature review for the Childwatch International Research Network*. Lismore: Southern Cross University, Centre for Children and Young People/Dunedin: University of Otago, Centre for Research on Children and Families. Retrieved from http://childethics.com/wp-content/uploads/2013/09/Powell-et-al-2012.pdf

Pyne, J. (2012). *Supporting gender independent children: A Rainbow Health Ontario fact sheet*. Toronto: Rainbow Health Ontario. Retrieved from https://www.rainbowhealthontario.ca/wp-content/uploads/woocommerce_uploads/2012/10/RHO_FactSheet_GIC_E1.pdf

Pyne, J. (2014). Health and wellbeing among gender independent children: A critical review of the literature. In E. Meyer & A. Pullen Sansfacon (Eds.), *Supporting transgender and gender creative youth: Schools, families, and communities in action* (pp. 26–40). New York: Peter Lang.

Reed, B., Rhodes, S., Schofield, P., & Wylie, K. (2009). *Gender variance in the UK. Prevalence, incidence, growth and geographic distribution*. London, UK: GIRES – the Gender Identity Research and Education Society. Retrieved from http://www.gires.org.uk/assets/Medpro-Assets/GenderVarianceUK-report.pdf

Sampson, R., & Laub, J. (2003). Desistance from crime over the life course. In J. Mortimer & M. Shanahan (Eds.), *Handbook of the life course* (pp. 295–309). New York: Springer.

Scheim, A. I., & Bauer, G. R. (2015). Sex and gender diversity among transgender persons in Ontario, Canada: Results from a respondent-driven sampling survey. *The Journal of Sex Research*, *52*(1), 1–14. Retrieved from https://doi.org/

10.1080/00224499.2014.893553. doi:10.1080/00224499.2014.893553

Schreier, H., & Ehrensaft, D. (2016, February 19). Want to know a child's gender? Ask. *San Francisco Gate*. Retrieved from http://www.sfgate.com/opinion/article/Want-to-know-a-child-s-gender-Ask-6843665.php

Schwartzapfel, B. (2013, March 14). *Born this way? The American Prospect*. Retrieved from: http://prospect.org/article/born-way

Serano, J. (2016, August 2). Desistance, detransition and disinformation: A guide for understanding transgender children debates. *Medium*. Retrieved from https://medium.com/@juliaserano/detransition-desistance-and-disinformation-a-guide-for-understanding-transgender-children-993b7342946e

Sherer, I. (2016). Social transition: Supporting our youngest transgender children. *Pediatrics*, *137*(3), e20154358. doi:10.1542/peds.2015-4358

Singh, A. (2013). Transgender youth of color and resilience: Negotiating oppression and finding support. *Sex Roles*, *68*, 690–702. doi:10.1007/s11199-012-0149-z

Steensma, T., Biemond, R., de Boer, F., & Cohen-Kettenis, P. (2011). Desisting and persisting gender dysphoria after childhood: A qualitative follow-up study. *Clinical Child Psychology & Psychiatry*, *16*(4), 499–516 doi:10.1177/1359104510378303

Steensma, T., & Cohen-Kettenis, P. (2011). Gender transition before puberty? *Archives of Sexual Behaviour*, *40*, 649–650. doi:10.1007/s10508-011-9752-2

Steensma, T., McGuire, J., Kreukels, B., Beekman, A., & Cohen-Kettenis, P. (2013). Factors associated with desistance and persistence of childhood gender dysphoria: A quantitative follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(6), 582–590. doi:10.1016/j.jaac.2013.03.016

Stouthamer-Loeber, M., Wei, E., Lober, R., & Masten, A. (2004). Desistance from persistent serious delinquency in the transition to adulthood. *Development & Psychopathology*, *16*(4), 897–918. doi:10.1017/S0954579404040064

Temple-Newhook, J., Winters, K., Pyne, J., Jamieson, A., Holmes, C., Feder, S., Picket, S., & Sinnott, M. (in press 2018). Teach your parents (and providers) well: A call for re-focus on the health of trans and gender-diverse children. *Canadian Family Physician*.

Turban, J. (2017). Transgender youth: The building evidence base for early social transition. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(2), 101–102. doi:10.1016/j.jaac.2016.11.008

Vanderburgh, R. (2009). Appropriate therapeutic care for families with pre-pubescent transgender/gender-dissonant children. *Child Adolescent Social Work Journal*, *26*, 135–154. doi:10.1007/s10560-008-0158-5

Wallace, R., & Russell, H. (2013). Attachment and shame in gender nonconforming children and their families: Toward a theoretical framework for evaluating clinical interventions. *International Journal of Transgenderism*, *14*(3), 113–126. doi:10.1080/15532739.2013.824845

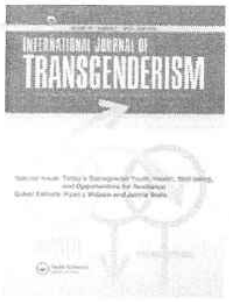
Wallien, M., & Cohen-Kettenis, P. (2008). Psychosexual outcome of gender-dysphoric children. *Journal of American Academy of Child & Adolescent Psychiatry*, *47*, 1413–1423. doi:10.1097/CHI.0b013e31818956b9

Winters, K. (2014, February). *Methodological questions in childhood gender identity 'desistance' research"*. Paper presented at the World Professional Association for Transgender Health Biennial Symposium, Bangkok, Thailand. Retrieved from https://gidreform.wordpress.com/2017/02/10/revisiting-flawed-research-behind-the-80-childhood-gender-dysphoria-desistance-myth/

Zucker, K., & Bradley, S. (1995). *Gender identity disorder and psychosexual problems in children and adolescents*. New York: Guilford Press.

Zucker, K. J., Wood, H., Singh, D., & Bradley, S. (2012). A developmental, biopsychosocial model for the treatment of children with gender identity disorder. *Journal of Homosexuality*, *59*, 369–397. doi:10.1080/00918369.2012.653309

Zuger, B. (1984). Early effeminate behavior in boys: Outcome and significance for homosexuality. *Journal of Nervous and Mental Disease*, *172*, 90–97. doi:10.1097/00005053-198402000-00005



International Journal of Transgenderism

ISSN: 1553-2739 (Print) 1434-4599 (Online) Journal homepage: https://www.tandfonline.com/loi/wijt20

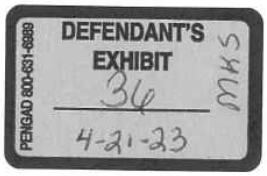
A critical commentary on "A critical commentary on follow-up studies and "desistence" theories about transgender and gender non-conforming children"

Thomas D. Steensma & Peggy T. Cohen-Kettenis

To cite this article: Thomas D. Steensma & Peggy T. Cohen-Kettenis (2018) A critical commentary on "A critical commentary on follow-up studies and "desistence" theories about transgender and gender non-conforming children", International Journal of Transgenderism, 19:2, 225-230, DOI: 10.1080/15532739.2018.1468292

To link to this article: https://doi.org/10.1080/15532739.2018.1468292

Published online: 29 May 2018.



INTERNATIONAL JOURNAL OF TRANSGENDERISM
2018, VOL. 19, NO. 2, 225–230
https://doi.org/10.1080/15532739.2018.1468292

A critical commentary on "A critical commentary on follow-up studies and "desistence" theories about transgender and gender non-conforming children"

Thomas D. Steensma and Peggy T. Cohen-Kettenis

Center of Expertise on Gender Dysphoria & Department of Medical Psychology, VU University Medical Center, Amsterdam, The Netherlands

KEYWORDS Desistence; gender dysphoria; gender incongruence; persistence; psychosexual development

The article entitled "A critical commentary on follow-up studies and "desistence" theories about transgender and gender non-conforming children" by Temple Newhook et al. (2018) is a plea to abandon longitudinal studies on the development of gender variant children as they do not respect children's autonomy. A few relatively recent studies are criticized and it is concluded that conducting longitudinal psychosexual outcome studies and acknowledging the children's feelings are contradictory. We agree that the longitudinal studies currently available have their limitations. We do, however, strongly disagree with the authors that studies on gender variant children's development should be abandoned and that our studies do not take children's needs and voices seriously or are unethical.

Before we address and discuss a number of specific criticisms in their paper, we first have two general, but crucial, remarks.

The authors claim in the very beginning of their paper that the 80% desistence rate of gender dysphoria (GD) is a number that is largely drawn on estimates from four follow-up studies: one from Canada (Drummond, Bradley, Peterson-Badali, & Zucker, 2008) and three from the Netherlands (Steensma, Biemond, de Boer, & Cohen-Kettenis, 2011; Steensma, McGuire, Kreukels, Beekman, & Cohen-Kettenis, 2013; Wallien & Cohen-Kettenis, 2008). Unfortunately, the authors do not seem to be entirely aware of the history behind the prevalence numbers. The first comprehensive and inclusive summary of historical follow-up studies on the psychosexual outcome in gender variant children was provided by Zucker and Bradley (1995). Later, these numbers were updated in Steensma et al. (2011) and updated again and further discussed in Ristori and Steensma (2016). Important to mention here is that in the calculation of the *overall* persistence rate in the literature the two studies by Steensma et al. (2011, 2013) were *never* used. Including the two Steensma et al. studies in the discussion about persistence rates by the authors (particularly in Table 1 in Temple Newhook et al. 2018) is in our view an odd choice and a methodologically incorrect one. The reason why both studies were not included is obvious: both studies did not aim to report on the prevalence of desistence or persistence of GD. The qualitative study in 2011, conducted among 25 participants, aimed to:

> obtain greater understanding of the processes and factors that may have contributed to the persistence and desistence of childhood gender dysphoria and sexual orientation development of gender dysphoric children (Steensma et al., 2011, p. 500).

As correctly indicated by the authors, this study offered some unique and valuable qualitative data. However, persistence rates of GD were never mentioned by Steensma et al. (2011). The quantitative study in 2013 aims were at the following:

> [to examine] possible factors associated with persistence of childhood GD by comparing a number of childhood variables (e.g., demographic background, GD, gender-variant behavior, psychological functioning, and quality of peer relations) between adolescent persisters and desisters who were clinically referred to our gender identity service in childhood. In addition to this, we examined psychosexual outcomes, body image, and the intensity of GD at the time of follow-up in adolescence. (Steensma et al., 2013, p. 583)

CONTACT Thomas D. Steensma t.steensma@vumc.nl VU University Medical Center, Center of Expertise on Gender Dysphoria, Department of Medical Psychology, P.O. box 7057, Amsterdam, 1007 MB, the Netherlands.

Again, because of the purpose and the design of this study we did not report prevalence numbers in the sample under study. Furthermore, the sample in the 2013 study did not include children in the younger age spectrum of the referred population to the Amsterdam clinic. Reporting prevalence of persistence and/or desistence in this sample would therefore not be reliable. The authors seem to have overlooked that both studies cannot be used "to support the 80% 'desistance' estimation."

Second, the two remaining studies come from two clinics with very different social contexts. The Toronto and Amsterdam clinical populations are not similar. This is illustrated clearly by a number of comparative studies on the social and emotional functioning of children and adolescents who are referred to the two clinics: in all studies psychological functioning in children and adolescents (measured through parental, teacher and/or self-report) showed to be better in Amsterdam than in Toronto (Cohen-Kettenis, Owen, Kaijser, & Bradley, 2003; de Vries, Steensma, VanderLaan, Cohen-Kettenis, & Zucker, 2016; Steensma et al., 2014). As noted in the discussions of these papers, we believe that differences in the social climate regarding gender variance may be an important factor underlying the differences. A sociometric study on peer group status (Wallien, Veentra, Kreukels, & Cohen-Kettenis, 2010) underscores this point. It was shown that there were no differences between gender-referred elementary school children and their classmates in victimization or bullying. Parents who come to our clinic are concerned about potential harm for their child if they would *not* acknowledge the child's gender experience. They are not focused on suppressing it, as the authors suggest by the sentence:

> Furthermore, even within Toronto and the Netherlands, this research was limited to children whose parents chose to bring them to a clinic for diagnosis and treatment and thus may have believed the child's difference was a problem, and one that required psychological treatment.

Our group is indeed a clinical one, but not all clinical groups are characterized by a suppressing attitude of their environment. We feel that, besides the unfounded assumptions that are made by the authors about the intentions of the supporting and caring parents we see in our clinic, the difference in social context is not properly taken into account in the paper, and that the conclusions that are drawn from our follow-up studies cast a very negative shadow on our clinical approach and the intentions of professionals in our clinic.

Methodological concerns

The authors claim to have identified strong methodological problems concerning the inclusion criteria for those children who were studied in childhood, the lack of acknowledgement of social context of the children, the age of the children at follow-up, and the misclassification of the adolescents who did not participate at the time of follow-up.

Although we do not believe that many of our non-responders are in fact persisters, we do agree with the authors that the persistence rates may increase in studies with different inclusion criteria. The classification of GD in the Wallien and Cohen-Kettenis (2008) study was indeed based on diagnostic criteria prior to DSM-5, with the possibility that some children were only gender variant in behavior. We have clearly described the characteristics of the included children (clinically referred and fulfilling childhood DSM criteria) and did not draw conclusions beyond this group, as has wrongly been done by others. The broadness of the earlier DSM criteria was also acknowledged by the American Psychiatric Association and World Health Organization. This was, among other things, a reason to tighten the diagnostic childhood criteria for DSM-5 and the proposed criteria for ICD-11. As we have stated elsewhere (Hembree et al., 2017; Steensma, 2013), we expect that future follow-up studies using the new diagnostic criteria may find higher persistence rates and hopefully shed more light on developmental routes of gender variant and transgender children.

When discussing the design of our studies (although the Wallien and Cohen-Kettenis (2008) is in fact the only one that should be discussed; see above), the authors question the actual persistence rate because of our classification of individuals who did not participate in the follow-up as desisters (the non-responders), and state that many (or all) of the non-responders at the time of follow-up may in fact be persisters. We agree that the current persistence rates may go up if a different methodology would be used (see above), but the suggested reasons why non-responders may actually be persisters are unlikely and farfetched. For instance, the authors suggested that children were lost at follow-up because they moved out of the country, were being treated elsewhere in the

Netherlands, or institutionalized. The chance that one of these situations occurred is, however, very low. The few families that had to go abroad for some time usually stayed in touch with the clinic, because in those days there were hardly any gender identity clinics for children and it was extremely difficult to get a prescription of puberty-blocking hormones. Besides language barriers and insurance problems, this would have been a major reason why going to clinics abroad out of dissatisfaction with the Dutch clinic would have been hard. Going to another Dutch clinic was not an option either. For a few years there was only one other child and adolescent clinic (which is now closed), but entrance to both clinics was centralized. So, if a child would reapply for treatment in adolescence in either one of the clinics, they would be recognized as a previously known child. Although it was not a favorable situation, there were—until fairly recently—few places families could go to if they wanted to avoid having contact with our clinic. If children were institutionalized and still wanted transgender care we usually collaborated with these institutions to see what could be done in terms of their medical treatment.

As suggested by the authors, it is possible that some of the adolescents classified as desisters at the time of follow-up were not following the straight transgender child–transgender adult path, but needed a much longer exploration time of their feelings before they decided they desired medical treatment. We were the first to report in the literature on children with GD who, in adolescence, indeed were not gender dysphoric to the extent that they wanted puberty blockers, gender affirming hormones, and surgery as soon as possible, but only came back to our clinic in adulthood (Steensma & Cohen-Kettenis, 2015). It is precisely this group that needs ample opportunity to explore the gender they feel comfortable with. For this reason, it would also be important to follow the development of transgender children over a much longer period than into young adulthood, instead of entirely refraining from developmental studies.

Theoretical concerns

The theoretical concerns brought up by the authors regarding the used terminology and the suggestion of an embedded pre-assumption in our studies that "stability of gender identity is a positive health outcome that should be prioritized for all children" may in our view be the result of a major difference in the operationalization of terms.

Unlike what is suggested, we have not studied the gender identities of the children. Instead we have studied the persistence and desistence of children's distress caused by the gender incongruence they experience to the point that they seek clinical assistance. As stated in our 2008 paper, we wanted to know more about the development of the children to find guidance in our clinical work.

> Clinically, it is also important to be able to discriminate between persisters and desisters before the start of puberty. If one was certain that a child belongs to the persisting group, interventions with gonadotropin-releasing hormone (GnRH) analogs to delay puberty could even start before puberty rather than after the first pubertal stages, as now often happens. (Wallien & Cohen-Kettenis, 2008, p. 1413)

We did so by looking at children's continuing (or discontinued) desire for medical gender affirming treatment when they entered puberty. This was a variable that could be relatively easy and reliably measured and was clinically highly relevant in the light of timing of puberty suppression decisions. Using the term desistence in this way does not imply anything about the identity of the desisters. The children could still be hesitating, searching, fluctuating, or exploring with regard to their gender experience and expression, and trying to figure out how they wanted to live. Apparently, they no longer desired some form of gender-affirming treatment at that point in their lives. The assumption that we considered all desisters as having a *fixed* cisgender identity is therefore an incorrect one.

We very much regret the assertion that we would only value certain "desisting outcomes" and the repeated suggestions by the authors that we would find cisgender outcomes "better" or "healthier." Whatever the outcome, our clinical approach has always aimed to work towards an outcome of children that makes a good quality of life possible, no matter what the outcome is. What is complicated, however, is that a continuously changing identity, although not considered inherently unhealthy or unfavorable, makes it hard to know at what moment what type of intervention may be in the best interest of the child, as some medical interventions are hard or impossible to reverse.

Ethical concerns

Before responding to the ethical concerns raised by the authors regarding the treatment and testing of children, their opinion about the goals of treatment that is provided, and the lack of consideration of children's autonomy in the studies, one should realize that choices that have been made in the past need to be seen in a historical perspective.

The literature on transgender and gender variant children is rapidly expanding. This is a positive development. Unfortunately, with the recent fast changing views and insights, historical developments are often overlooked. With regard to the situation in the Netherlands, one can say that 30 years ago Dutch children who were struggling with gender identity issues and who were unhappy about their assigned gender were on nobody's agenda. At the time, they were occasionally seen by psychotherapists and psychiatrists and received various forms of therapy, very much depending on the theoretical or personal views of their therapist.

In the late 1980s families started to come to our clinic with pre-pubertal children, probably because they had heard that we were counseling (and later treating) transgender adolescents. They usually came with children aged 6 years or older (Cohen-Kettenis et al., 2003), asking for pedagogical advice. Behavioral or other problems (e.g., anxiety, depression, ADHD, or ODD) were another reason for seeking clinical support. Parents wanted to know whether these problems and the gender variance/GD of the children were related, and they preferred counseling or treatment for the general problems of the child by a clinic with experience in the field of transgender health.

As one of the first clinics for transgender children, we needed to develop our own approach. In order to provide personalized care, we not only needed information on the gender identity preferences, behaviors, and interests, but also about the general, psychological, and social functioning of the individual child. Obviously, an impulsive, oppositional child with ADHD and poor verbal skills needs a very different approach than an eloquent, withdrawn, and anxious child. Gathering such information (partly by psychological testing, which was carried out, according to the child psychology and child psychiatry standards, in a proper, safe, and child-friendly way and environment) provides useful information about the functioning of the child. Besides the fact that psychological testing is a fundamental part of psychological assessment in (child) psychology (Cronbach, 1990; Saklofske, Reynolds, & Schwean, 2013), such an approach is efficient, useful, *and* ethical, also in case of children who have already socially transitioned. In our view, it has little to do with thwarting the child's gender experience.

Besides needing information on individual children's functioning, we also wanted to have a better idea about developmental trajectories of transgender children, their specific vulnerabilities that might need attention, and/or a clinical approach that could be beneficial for the children. As we were the first (in the world) to provide adolescents with puberty-blocking treatment, it was important for us to know more about the lowest age for responsibly starting with this treatment; one would not want to risk adverse effects by prescribing medication to children who would not benefit from it, and giving such medication is not something one would want to start and stop repeatedly. Longitudinal studies in children gave us some insight in the stability of the desire for treatment in adolescence and should therefore be seen as valuable in our field instead of ethically wrong or useless. Children were never involved in these studies against their will, as is suggested.

Interpretive concerns

The authors question our clinical approach which was (partly) relying on the findings of our own and other studies, in particular with regard to social transitioning in early childhood.

As described elsewhere by others (Byne et al., 2012; Drescher, 2013), the clinical approach in the counseling of pre-pubertal children in the Netherlands has always been primarily focusing on treating the co-existing problems and only monitoring the gender development of the children. However, such a summary is too limited and incomplete. Besides interventions that focus on the co-existing problems of the child and/or the family, we focused, among other things, on helping parents and the child to bear the uncertainty of the child's psychosexual outcome, provide psychoeducation to help the child and the family to make balanced decisions regarding topics such as the child's coming out or early social transitioning. In these discussions the child always had an important voice in the decisions that were made. We encouraged

parents to support the child's gender experience, while at the same time helping the child to deal with complex issues in their lives. However, the chance that the desire for a medical transition may desist in the future was a topic of discussion as well and something that had to be included in the decisions. Parents were encouraged to provide enough space for their child to explore their gender dysphoric feelings, while at the same time keeping all future outcomes open (e.g., de Vries & Cohen-Kettenis, 2012; Ristori & Steensma, 2016).

With regard to the topic of social transitioning, one should bear in mind that before the year 2000 we only saw one pre-pubertal child who desired a social transition and the number only slowly increased in the next few years (Steensma & Cohen-Kettenis, 2011). By listening closely to the children and their families, these children were not frustrated in their desire. Instead we gave support as well as advice based on what we knew about these children as a group. We certainly did not advertise that transitioning was good for all gender variant children, but stressed that such decisions should be personalized. We felt that the safest modus was to leave all possibilities open, irrespective of transitioning. For most children the situation changed dramatically when they approached puberty. Some were no longer interested in anything related to gender variance, some remained hesitant, whereas others intensified their contact with the clinic because they desired to proceed with medical interventions. As a high percentage did not stay with us for medical care we were quite careful not to impose any (gender) pathway. Recently the number of children who already had transitioned at their first visit to the clinic has exploded (Steensma & Cohen-Kettenis, 2011). For these families the question of whether or not the child should transition is obsolete.

Conclusion

Having responded to the many comments of Temple Newhook and colleagues we want to stress that we do not consider the methodology used in our studies as optimal (as previously indicated and discussed by ourselves in Ristori & Steensma, 2016), or that the terminology used in our communications is always ideal. As shown, it may lead to confusion and wrong inferences. We also agree that the persistence/desistence terms suggest or even induce binary thinking. In the last few years many terms in our field have changed. Evidently, we need to look for better terms covering the various possible outcomes and better indicating possible fluidity in the "desisting" group.

We, however, very much regret that our careful search for optimal care of the gender variant child by trying to gain more insight in the development of this group is taken for unethical behavior. Instead of polarization and accusations that can be read so often in the literature, collaboration in gathering more (and better) information about the development of gender variant children in different social contexts will better serve the quality of life of transgender children. Children and their families do not need arguing clinicians, but responsible care that is based on good evidence.

Declaration of conflict of interest

The authors have no conflict of interest to declare.

References

Byne, W., Bradley, S. J., Coleman, E., Eyler, A. E., Green, R., Menvielle, E. J., … Tompkins, D. A. (2012). Report of the American psychiatric association task force on treatment of gender identity disorder. *Archives of Sexual Behavior, 41*, 759–796. doi:10.1007/s10508-012-9975-x

Cohen-Kettenis, P. T., Owen, A., Kaijser, V. G., Bradley, S. J., & Zucker, K. J. (2003). Demographic characteristics, social competence, and problem behavior in children with gender identity disorder: A cross-national, cross-clinic comparative analysis. *Journal of Abnormal Child Psychology, 31*, 41–53. doi:10.1023/A:1021769215342

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York, NY: Harper & Row

de Vries, A. L., & Cohen-Kettenis, P. T. (2012). Clinical management of gender dysphoria in children and adolescents: The Dutch approach. *Journal of Homosexuality, 59*, 301–320. doi:10.1080/00918369.2012.653300

de Vries, A. L., Steensma, T. D., VanderLaan, D. P., Cohen-Kettenis, P. T., & Zucker, K. J. (2016). Poor peer relations predict parent- and self-reported behavioral and emotional problems of adolescents with gender dysphoria: A cross-national, cross-clinic comparative analysis. *European Child and Adolescent Psychiatry, 25*, 579–588. doi:10.1007/s00787-015-0764-7

Drescher, J. (2013). Controversies in gender diagnoses. *LGBT Health, 1*, 10–14. doi:10.1089/lgbt.2013.1500

Drummond, K. D., Bradley, S. J., Peterson-Badali, M., & Zucker, K. J. (2008). A follow-up study of girls with gender identity disorder. *Developmental Psychology, 44*, 34–45. doi:10.1037/0012-1649.44.1.34

Hembree, W. C., Cohen-Kettenis, P. T., Gooren, L., Hannema, S. E., Meyer, W. J., Murad, M. H., … T'Sjoen, G. (2017). Endocrine treatment of gender-dysphoric/gender-incongruent persons: An endocrine society clinical practice guideline. *Journal of Clinical Endocrinology & Metabolism, 102*, 3869–3903. doi:10.1210/jc.2017-01658

Ristori, J., & Steensma, T. D. (2016). Gender dysphoria in childhood. *International Review of Psychiatry, 28*, 13–20 doi:10.3109/09540261.2015.1115754

Saklofske, D. H., Reynolds, C. R., & Schwean, V. L. (2013). *The Oxford handbook of child psychological assessment*. New York, NY: Oxford University Press.

Steensma, T. D. (2013). *From gender variance to gender dysphoria: Psychosexual development of gender atypical children and adolescents* (Dissertation). VU University, Amsterdam, the Netherlands:

Steensma, T. D., Biemond, R., de Boer, F., & Cohen-Kettenis, P. T. (2011). Desisting and persisting gender dysphoria after childhood: A qualitative follow-up study. *Clinical Child Psychology and Psychiatry, 16*, 499–516. doi:10.1177/1359104510378303

Steensma, T. D., & Cohen-Kettenis, P. T. (2011). Gender transitioning before puberty? *Archives of Sexual Behavior, 40*, 649–650. doi:10.1007/s10508-011-9752-2

Steensma, T. D., & Cohen-Kettenis, P. T. (2015). More than two developmental pathways in children with gender dysphoria? *Journal of the American Academy of Child and Adolescent Psychiatry, 54*, 147–148. doi:10.1016/j.jaac.2014.10.016

Steensma, T. D., McGuire, J. K., Kreukels, B. P., Beekman, A. J., & Cohen-Kettenis, P. T. (2013). Factors associated with desistence and persistence of childhood gender dysphoria: A quantitative follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*, 582–590. doi:10.1016/j.jaac.2013.03.016

Steensma, T. D., Zucker, K. J., Kreukels, B. P. C., VanderLaan, D. P., Wood, H., Fuentes, A., & Cohen-Kettenis, P. T. (2014). Behavioral and emotional problems on the teacher's report form: A cross-national, cross-clinic comparative analysis of gender dysphoric children and adolescents. *Journal of Abnormal Child Psychology, 42*, 635–647. doi:10.1007/s10802-013-9804-2

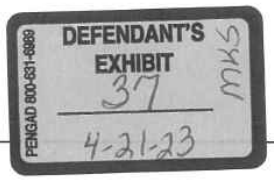
Temple Newhook, J., Pyne, J., Winters, K., Feder, S., Holmes, C., Tosh, J., Sinnott, M., Jamieson, A., & Picket, S. (2018). A critical commentary on follow-up studies and "desistance" theories about transgender and gender non-conforming children. *International Journal of Transgenderism*. Advance online publication. doi:10.1080/15532739.2018.1456390.

Wallien, M. S., & Cohen-Kettenis, P. T. (2008). Psychosexual outcome of gender-dysphoric children. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*, 1413–1423. doi:10.1097/CHI.0b013e31818956b9

Wallien, M. S. C., Veentra, R., Kreukels, B. P. C., & Cohen-Kettenis, P. T. (2010). Peer Group status of gender dysphoric children: A sociometric study. *Archives of Sexual Behavior, 39*, 553–560. doi:10.1007/s10508-009-9517-3

Zucker, K. J., & Bradley, S. (1995). *Gender identity disorder and psychosexual problems in children and adolescents*. New York, NY: Guilford Press.

THE JOURNAL OF
SEXUAL MEDICINE



ORIGINAL RESEARCH

TRANSGENDER HEALTH

The Amsterdam Cohort of Gender Dysphoria Study (1972–2015): Trends in Prevalence, Treatment, and Regrets

Chantal M. Wiepjes,[1,2] Nienke M. Nota,[1,2] Christel J. M. de Blok,[1,2] Maartje Klaver,[1,2] Annelou L. C. de Vries,[2,3] S. Annelijn Wensing-Kruger,[2,4] Renate T. de Jongh,[1] Mark-Bram Bouman,[2,5] Thomas D. Steensma,[2,4] Peggy Cohen-Kettenis,[2,4] Louis J. G. Gooren,[1,2] Baudewijntje P. C. Kreukels,[2,4] and Martin den Heijer, MD, PhD[1,2]

ABSTRACT

Background: Over the past decade, the number of people referred to gender identity clinics has rapidly increased. This raises several questions, especially concerning the frequency of performing gender-affirming treatments with irreversible effects and regret from such interventions.

Aim: To study the current prevalence of gender dysphoria, how frequently gender-affirming treatments are performed, and the number of people experiencing regret of this treatment.

Methods: The medical files of all people who attended our gender identity clinic from 1972 to 2015 were reviewed retrospectively.

Outcomes: The number of (and change in) people who applied for transgender health care, the percentage of people starting with gender-affirming hormonal treatment (HT), the estimated prevalence of transgender people receiving gender-affirming treatment, the percentage of people who underwent gonadectomy, and the percentage of people who regretted gonadectomy, specified separately for each year.

Results: 6,793 people (4,432 birth-assigned male, 2,361 birth-assigned female) visited our gender identity clinic from 1972 through 2015. The number of people assessed per year increased 20-fold from 34 in 1980 to 686 in 2015. The estimated prevalence in the Netherlands in 2015 was 1:3,800 for men (transwomen) and 1:5,200 for women (transmen). The percentage of people who started HT within 5 years after the 1st visit decreased over time, with almost 90% in 1980 to 65% in 2010. The percentage of people who underwent gonadectomy within 5 years after starting HT remained stable over time (74.7% of transwomen and 83.8% of transmen). Only 0.6% of transwomen and 0.3% of transmen who underwent gonadectomy were identified as experiencing regret.

Clinical Implications: Because the transgender population is growing, a larger availability of transgender health care is needed. Other health care providers should familiarize themselves with transgender health care, because HT can influence diseases and interact with medication. Because not all people apply for the classic treatment approach, special attention should be given to those who choose less common forms of treatment.

Strengths and Limitations: This study was performed in the largest Dutch gender identity clinic, which treats more than 95% of the transgender population in the Netherlands. Because of the retrospective design, some data could be missing.

Conclusion: The number of people with gender identity issues seeking professional help increased dramatically in recent decades. The percentage of people who regretted gonadectomy remained small and did not show a tendency to increase. **Wiepjes CM, Nota NM, de Blok CJM, et al. The Amsterdam Cohort of Gender Dysphoria Study (1972–2015): Trends in Prevalence, Treatment, and Regrets. J Sex Med 2018;15:582–590.**

Key Words: Transgender; Prevalence; Regret; Gender-Affirming Hormones; Gender-Affirming Surgery

Received December 8, 2017. Accepted January 28, 2018.

[1]Department of Endocrinology, VU University Medical Center, Amsterdam, the Netherlands;

[2]Center of Expertise on Gender Dysphoria, VU University Medical Center, Amsterdam, the Netherlands;

[3]Department of Child and Adolescent Psychiatry, VU University Medical Center, Amsterdam, the Netherlands;

[4]Department of Medical Psychology, VU University Medical Center, Amsterdam, the Netherlands;

[5]Department of Plastic, Reconstructive, and Hand Surgery, VU University Medical Center, Amsterdam, the Netherlands

https://doi.org/10.1016/j.jsxm.2018.01.016

INTRODUCTION

Gender dysphoria (GD) refers to the distress related to a marked incongruence between one's assigned sex at birth and the experienced gender later in life.[1] In this study, we define transwomen as having a male birth assignment and transmen as having a female birth assignment who might receive medical treatment to adapt their physical characteristics to their experienced gender. This treatment can include puberty suppression (PS), gender-affirming hormonal treatment (HT), and gender-affirming surgery.

It has been widely observed that the transgender population is growing and broadening.[2,3] This increase in the transgender population raises several questions, especially concerning the frequency of performing gender-affirming treatments with irreversible effects and regret from such interventions.

There are no reliable estimations of the current prevalence of transgender people who actually have received gender-affirming treatment (including HT), because most recent studies are based on questionnaires[4,5] or data about gender-affirming surgery only.[6,7] In most countries transgender care is performed by multiple health care providers (eg, university clinics or general practitioners), which makes it difficult to provide these numbers. In contrast, in the Netherlands, more than 95% of the transgender population has received treatment in only 1 center, the gender identity clinic at the VU University Medical Center (VUmc; Amsterdam, the Netherlands), currently known as the Center of Expertise on Gender Dysphoria.[8–10] This center started treating adults in 1972. From 1987 to 2002, children and adolescents were seen by a mental health specialist in the Utrecht University Medical Center (Utrecht, the Netherlands). After they were considered eligible, they could receive medical treatment in the VUmc, which consisted of PS (usually by gonadotropin-releasing hormone analogues), followed by HT (see Kreukels and Cohen-Kettenis[11] for the treatment protocol for adolescents diagnosed with GD). After 2002, the Utrecht clinic stopped seeing adolescents and the diagnostics were performed in the VUmc. Adult people are referred to a psychologist or psychiatrist for the diagnostic phase after an initial screening. People diagnosed with GD can start HT if they are considered eligible. HT consists of testosterone for transmen and estrogens, often combined with antiandrogens, for transwomen. In the 1st year of HT, checkups are performed every 3 months. After a minimum of 12 months of HT, gender-affirming surgery can be performed, including mastectomy and hysterectomy with oophorectomy in transmen and breast augmentation and vaginoplasty (including orchiectomy) in transwomen. After gonadectomy (oophorectomy or orchiectomy), people are usually seen every 1 to 2 years for clinical follow-up.

In the present study we included the complete population seen at the gender identity clinic of the VUmc from 1972 through December 2015 to assess the current prevalence of transgender people who received medical treatment, the frequency of specific medical treatments performed, and the numbers of people who received HT in line with their sex assigned at birth because they regretted undergoing gonadectomy.

METHODS

Study Design and Patient Selection

After approval of the local ethics committee, a retrospective medical record review was performed to identify all people seen in our gender identity clinic from 1972 until December 2015. Data were collected from the hospital registries of the VUmc. The total study population was defined as people who had been diagnosed with 1 of the following *International Classification of Diseases* diagnoses: 302.5 (transsexualism), 302.6 (gender identity disorder not otherwise specified), or 302.85 (gender identity disorder in adolescent or adult) according to the 9th edition or F64 (gender identity disorders) according to the 10th edition.[12] In addition, the administrative employees of our gender identity clinic registered everyone who was referred to our gender identity clinic since the early 1970s. People reported on this list also were included in the study population. Some people of this study population have been described in previous studies.[9,13–18] People were excluded from the study if they had been registered at our gender identity clinic but had actually never visited the clinic or if they had presented with other complaints than gender identity issues. Because of the retrospective design and the large study population, necessity for informed consent was waived by our local ethics committee.

Hospital Registries

The hospital registries store clinical data obtained during regular patient care performed in our center, including medical diagnoses (since 1985), medication prescriptions (since 2000), surgical interventions (since 2006), laboratory test results (since 2004), radiology results (since 1993), and visit dates (since 2007). The 1st visit was defined as the 1st appointment with the psychologist, psychiatrist, pediatrician, endocrinologist, or gynecologist for health care related to gender identity.

Clinical Data Collection

Not all data were available from the hospital registries, particularly older data or surgeries performed in other centers. To generate the most reliable results, the medical records of all people who composed the study population were checked. All people were classified as transwomen or transmen (based on the sex assigned at birth), and date of birth and death were noted. The following categories were included: the individual was in the diagnostic stage, the individual did not start HT, or the individual was on HT. Start of HT was defined as the 1st date gender-affirming hormones were prescribed by a physician in our gender identity clinic after a confirmed GD diagnosis, irrespective of previous gender-affirming hormone use. Of the people who started HT, baseline and follow-up data, including

1st visit, medical history, medication use, prior gender-affirming hormone use, start date and type of PS and HT, and date of gonadectomy, were collected. Some people regretted the interventions they had undergone. Transwomen who started testosterone treatment after vaginoplasty or transmen who started estrogen treatment after oophorectomy and expressed regret were categorized as those who experienced regret. Reasons for regret as reported in their medical records were noted. Dates were set to the 1st of the month and personal identification data were removed from the research database.

Statistical Analysis

The total number of people visiting the clinic each year and their median age were reported separately for transwomen and transmen and were stratified for age at the 1st visit: children were younger than 12 years, adolescents were 12 to 18 years old, and adults were at least 18 years old. The percentage of people who started HT within 5 years after the 1st visit was reported for each year. The prevalence was calculated for people at least 12, at least 16, 12 to 18, 18 to 30, 30 to 50, and at least 50 years old by using the total number of people in these age groups who received medical treatment in our center until 2015, excluding deceased people. The total populations of these age groups in the Netherlands in 2015 were provided by the Central Bureau of Statistics of the Netherlands. The percentage of people who underwent gonadectomy within 5 years after starting HT was reported. For calculation of the total percentage of the study population who had undergone gonadectomy, only people at least 18 years old who used HT for at least 1.5 years were included, because these were requirements for surgery. People who regretted their medical transition are reported as the percentage of the total population of transwomen and transmen who underwent gonadectomy. In adults, time from 1st visit to start of HT or gonadectomy, if applicable, are expressed as median days with interquartile range (IQR). Total follow-up time was calculated for every individual who started HT and was expressed as years from the 1st visit to the last visit. Prevalence with 95% CI was calculated using OpenEpi.[19] All other analyses were performed using STATA 13.1 (StataCorp, College Station, TX, USA).

RESULTS

1st Visit

6,793 people presented for gender-affirming treatment, with more transwomen (65.2%) than transmen (34.8%; Table 1). The number of people attending the gender identity clinic increased over time (Table 2), whereas the median age of adults at the time of their 1st visit decreased (Figure 1). The median age at the 1st visit was younger for adult transmen (25 years; IQR = 21–35 years) than for adult transwomen (33 years; IQR = 25–42 years). Although historically more transwomen than transmen presented for treatment, more transmen than

Table 1. Treatment patterns of total study population, stratified for age groups and for transwomen and transmen*

	Transwomen	Transmen	Total	Ratio of transwomen to transmen
Total study population, N (%)	4,432 (65.2)	2,361 (34.8)	6,793 (100)	1.9:1
Adults (≥18 y)	3,809	1,624	5,433	2.3:1
Age (y)†, median (IQR; max)	33 (25–42; 81)	25 (21–35; 73)	31 (23–41; 81)	
Started HT§, %	68.9	72.9	69.9	
Underwent gonadectomy‖, %	75.3	83.8	77.7	
Adolescents (12–18 y)	330	482	812	0.7:1
Age (y)†, median (IQR)	16 (15–17)	16 (15–17)	16 (15–17)	
Started PS‡, %	28.7	50.8	41.0	
Stopped PS, %	4.1	0.7	1.9	
Started HT§ without PS, %	33.9	30.8	32.2	
Underwent gonadectomy‖, %	79.5	77.2	78.2	
Children (<12 y)	293	255	548	1.1:1
Age (y)†, median (IQR)	8 (7–10)	9 (8–11)	9 (7–10)	
Started PS‡,¶, %	33.6	49.1	40.3	
Regret#, % (n)	0.6 (11)	0.3 (3)	0.5 (14)	2.0:1

HT = gender-affirming hormonal therapy; IQR = interquartile range; max = maximum; PS = puberty suppression.
*From 1987 through 2002, children and adolescents were seen at the Utrecht University Medical Center and then at the VU University Medical Center only if they could begin medical treatment.
†Age is defined as the age at the 1st visit to the VU University Medical Center, Amsterdam.
‡Only those who reached the age of eligibility (usually ≥12 years old) could undergo PS.
§Only in people at least 16 years old.
‖Only people treated with gender-affirming hormones for at least 1.5 years and at least 18 years old (orchiectomy in transwomen and oophorectomy in transmen).
¶Those who were too old (≥18 years) after the diagnostic phase for PS could begin directly with HT.
#Only those people who underwent gonadectomy.

Table 2. Description of adult study population for every 5-year cohort

	1st visit, n	Started HT*, %	Age (y) at start of HT, median (IQR)	Previous HT, %	Underwent gonadectomy†, %
Transwomen (≥18 y)					
1972–1979	119	89.9	33 (26–40)	16.8	79.4
1980–1984	189	88.4	33 (25–40)	12.6	71.9
1985–1989	319	75.9	31 (25–39)	15.3	76.5
1990–1994	392	65.8	30 (25–41)	20.5	76.7
1995–1999	522	65.5	34 (27–41)	26.6	78.7
2000–2004	605	56.0	38 (30–45)	29.2	67.3
2005–2009	476	61.6	39 (29–47)	22.9	68.6
2010–2014	926 (138‡)	60.9‡	32 (23–42)‡	29.8‡	NA
Transmen (≥18 y)					
1972–1979	30	96.7	24 (21–30)	10.3	72.4
1980–1984	69	84.1	24 (21–32)	3.5	82.8
1985–1989	105	84.8	24 (21–30)	1.1	79.8
1990–1994	142	69.0	27 (21–33)	7.1	88.8
1995–1999	177	65.0	29 (24–37)	7.0	88.7
2000–2004	207	63.3	32 (26–39)	5.3	87.0
2005–2009	185	63.8	29 (23–37)	3.4	81.4
2010–2014	518 (70‡)	71.4‡	24 (21–37)‡	0‡	NA

HT = gender-affirming hormonal therapy; IQR = interquartile range; NA = not applicable.
*People who started HT within 5 years after the 1st visit.
†People who had this procedure within 5 years after the start of HT.
‡Only in people who had their 1st visit 5 years before December 31, 2015 (n = 138 transwomen; n = 70 transmen).

transwomen applied for treatment in 2015. This change in sex ratio was mainly due to the increase in adolescent transgender boys, because the ratio of transwomen to transmen in adults remained stable over time.

Prevalence and Treatment

At the end of 2015, 3,838 transgender people at least 16 years old had received medical treatment and were not deceased. Because the total population of people at least 16 years old in the Netherlands in 2015 was 13,870,426, the prevalence was 27.7 per 100,000 people (95% CI = 26.8–28.6), or 1:3,600. Stratification for transwomen and transmen showed a prevalence of 36.4 (95% CI = 35.0–37.8) per 100,000 people (or 1:2,800) for men (transwomen) and 19.3 (95% CI = 18.3–20.3) per 100,000 people (or 1:5,200) for women (transmen). The calculation of prevalence numbers of people at least 12 years old and specific age groups are presented in Table 3.

The percentage of adult people who started HT within 5 years after the 1st visit decreased over time, whereas the percentage of people who underwent gonadectomy within 5 years after starting HT remained stable (Figure 2). Of the total study population at least 18 years old treated with HT for at least 1.5 years, 75.6% of transwomen (n = 1,742) and 82.4% of transmen (n = 885) underwent gonadectomy. The median time from the 1st visit to the start of HT for adults was 327 days (IQR = 36–570 days) and from the 1st visit to gonadectomy was 1,029 days (IQR = 679–1,465 days). The median follow-up time for people treated with HT was 6.4 years (range = 0.4–41.6 years).

Of adolescents, 41.0% started PS, whereas only 1.9% of these adolescents stopped PS and did not start HT (Table 1). 32.2% of adolescents started directly with HT, because they were too old (≥18 years) to start with PS after the diagnostic phase.

Regret

Regret was identified in 0.6% of transwomen and 0.3% of transmen who underwent gonadectomy. The characteristics of these people are presented in Table 4. Their ages at start of HT ranged from 25 to 54 years, and they expressed their regrets 46 to 271 months after initiation of HT. Reasons for regret were divided into social regret, true regret, or feeling non-binary. Transwomen who were classified as having social regret still identified as women, but reported reasons such as "ignored by surroundings" or "the loss of relatives is a large sacrifice" for returning to the male role. People who were classified as having true regret reported that they thought gender-affirming treatment would be a "solution" for, for example, homosexuality or personal acceptance, but, in retrospect, regretted the diagnosis and treatment.

DISCUSSION

The aim of this study was to generate a dataset of all individuals who presented to our clinic for gender-affirming care from 1972 to 2015. We found that the number of people with

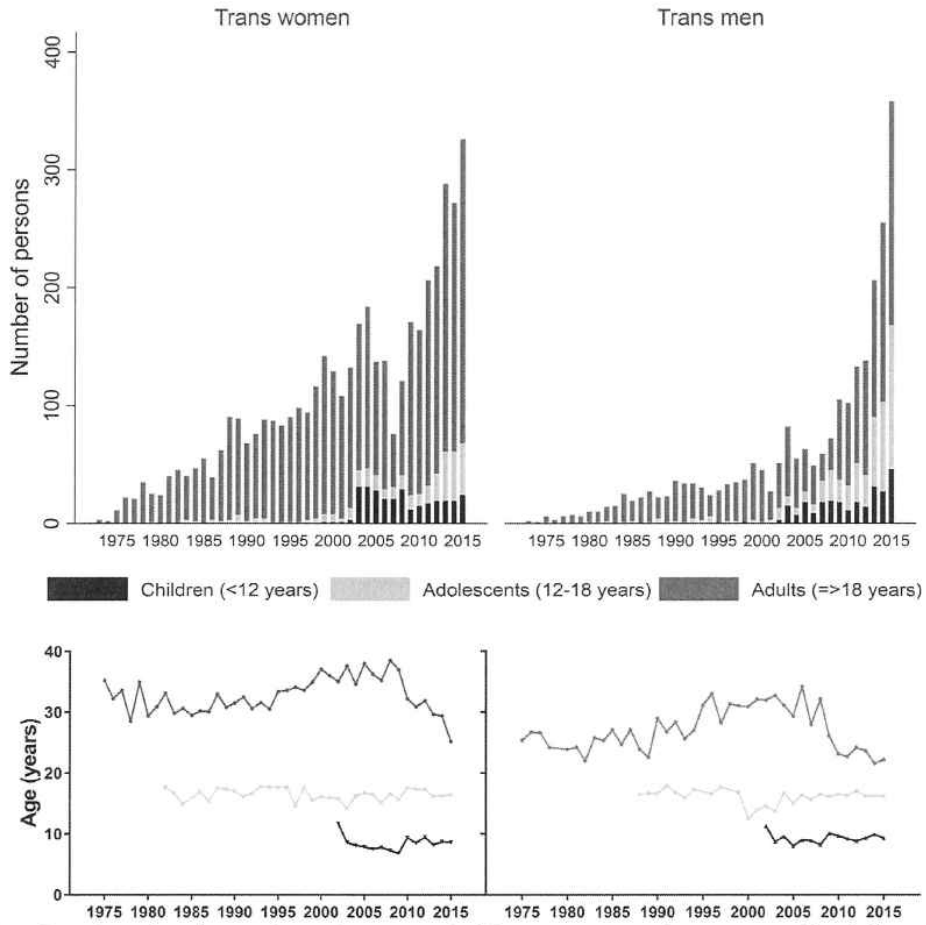


Figure 1. Number of people and median age for each year, stratified for transwomen and transmen and for children (<12 years), adolescents (12–18 years), and adults (≥18 years). Age is defined as age at the 1st visit to the VU University Medical Center, Amsterdam. From 1987 through 2002, children and adolescents were seen at the Utrecht University Medical Center and then at the VU University Medical Center only if they could begin medical treatment.

gender identity issues who sought professional help increased dramatically in recent decades and that the median age of adults at presentation decreased. The ratio of transwomen to transmen remained stable over the years for adults, whereas in adolescents the population of transgender boys increased compared with the population of transgender girls. Currently, more transgender boys than transgender girls are seen. This phenomenon also has been described by Aitken et al.[17] The age at the 1st visit was

Table 3. Prevalence numbers, specified for different age groups*

	Total population		Male sex assigned at birth (transwomen)		Female sex assigned at birth (transmen)	
Age (y)	Per 100,000	1 per	Per 100,000	1 per	Per 100,000	1 per
≥12	26.9 (26.1–27.8)	3,700	34.8 (33.5–36.2)	2,900	19.3 (18.3–20.3)	5,200
≥16	27.7 (26.8–28.6)	3,600	36.4 (35.0–37.8)	2,800	19.3 (18.3–20.3)	5,200
12–18	16.0 (13.9–18.4)	6,300	11.1 (8.8–14.1)	9,000	21.0 (17.7–25.1)	4,800
18–30	35.7 (33.5–38.2)	2,800	30.3 (27.4–33.4)	3,300	41.4 (37.9–45.1)	2,400
30–50	30.5 (29.0–32.2)	3,300	40.1 (37.6–42.8)	2,500	21.0 (19.2–23.0)	4,800
≥50	23.0 (21.9–24.2)	4,300	37.6 (35.5–39.8)	2,700	9.7 (8.7–10.8)	10,300

*Data are presented as number (95% CI).

Table 4. Characteristics of people with regret

Case	Type	Year started HT	Age (y) at start of HT	Year of gonadectomy	Time after HT (mo)	Time after gonadectomy (mo)	Reversal surgery	Reason for regret
1	M-F-M	1978	31	1979	±153	±130	None	Social acceptance
2	M-F-M	1982	25	1984	±54	±27	Mastectomy	Social acceptance
3	M-F-M	1986	47	1988	±216	±197	Mastectomy	Social acceptance
4	M-F-M	1988	33	1990	±186	±167	None	True regret
5	M-F-M	1988	38	1990	±70	±44	Mastectomy	Social acceptance
6	M-F-M	1991	41	1993	±67	±49	Mastectomy, vaginectomy, phalloplasty	Social acceptance
7	M-F-M	1991	38	1995	±271	±225	Mastectomy	True regret
8	M-F-M	1993	30	1994	±79	±61	None	Feels non-binary
9	M-F-M	1996	33	1997	±90	±73	Mastectomy, phalloplasty	True regret
10	M-F-M	1997	43	1999	±46	±27	Mastectomy	True regret
11	M-F-M	2004	54	2007	±130	±92	Mastectomy, vaginectomy	True regret
12	F-M-F	1987	25	1990	±91	±50	Breast augmentation, remove testicular implants	True regret
13	F-M-F	1990	34	1993	±102	±74	Remove testicular implants	Feels non-binary
14	F-M-F	1993	31	1997	±258	±212	None	True regret

F-M-F = female to male to female; HT = hormonal treatment; M-F-M = male to female to male.

older for adult transwomen than for transmen. The percentage of adult people starting HT within 5 years after the 1st visit decreased over time, whereas the percentage of people who underwent gonadectomy within 5 years after starting HT remained stable. Of the total population treated with HT, 77.8% underwent a gonadectomy. Only a very small percentage of people who underwent gonadectomy regretted their decision, expressed as the start of HT in line with their sex assigned at birth.

An explanation for the increase in referrals could be the increased attention in society and media, which contributes not only to awareness of the existence of GD and possibilities for medical treatment but also to greater social acceptance. In addition, information about transgender identities has become much more accessible through the internet within the past decade, which could lead to an earlier recognition of gender identity issues. Also, transgender and gender non-binary individuals might be more willing to access care and more access to care has become available.

The increase in the prevalence of people with GD who sought medical treatment in the Netherlands (1:11,900 transwomen and 1:30,400 transmen in 1990[8] vs 1:2,800 transwomen and 1:5,200 transmen currently) suggests that the transgender population is dramatically increasing. The highest prevalence for transwomen was found for the 30- to 50-year age group (1:2,500), whereas that for transmen was found in the 18- to 30-year age group (1:2,400). Transgender people in the Netherlands seem to experience a reasonable degree of acceptance owing to a tolerant social climate in contrast to many other countries.[20] For example, medical costs are reimbursed by medical insurance companies, and it is possible to change the legal sex status (even without gonadectomy). These points can lead to a lower threshold to seek help, making this study population useful for an adequate estimation of the current prevalence of people with GD who seek medical treatment. More than 95% of transgender people are treated in our gender identity clinic. However, not all transgender people seek medical help. Some use self-medication or go abroad for treatment. Therefore, these numbers might still be an underestimation of the real prevalence. Our data represent a population that actively sought help in a medical setting. In 2012, a Dutch study of non-clinical people reported that 0.6% (1:167) of those with male sex assigned at birth and 0.2% (1:500) of those with female sex assigned at birth reported an incongruent gender identity with a wish for hormones or surgery.[21] However, that was a population-based study with a response rate of 20.9%, which could lead to non-response bias. In addition, the existence of incongruent gender identities was based on self-report and no detailed assessment of GD was performed, which could have led to higher prevalence rates.

An interesting finding is the percentage of children who were referred in childhood (before 12 years of age) and who started PS when the GD persisted and the eligibility criteria were fulfilled. This 40% of children who started PS is almost identical to the 39% of persistence of childhood GD reported in a previous

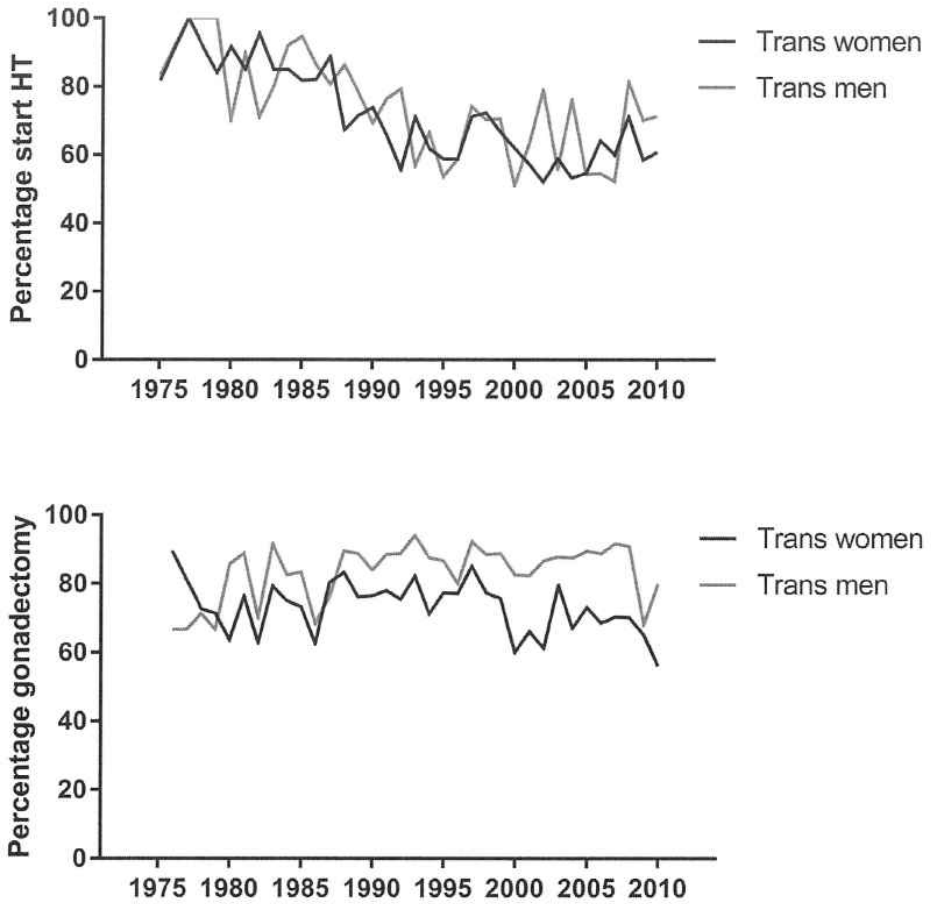


Figure 2. Top panel shows percentage of transgender adults beginning gender-affirming HT within 5 years after the 1st visit, stratified for transwomen and transmen. Bottom panel shows percentage of transgender adults with occurrence of gonadectomy within 5 years after starting HT for each year, stratified for transwomen and transmen. Year is defined as the year of the 1st visit. HT = gender-affirming hormonal treatment.

Dutch study (using a smaller cohort of children).[22] In addition, the finding that the persistence is higher in natal girls (49.1%) compared with natal boys (33.6%) is in line with observations in previous follow-up studies on the persistence of GD in children (for an overview, see Ristori and Steensma[23]).

Remarkably, we found a decrease over time in the percentage of referred adult people who actually started HT. This finding might be explained by the fact that in the past it was harder to find information about GD and its treatment, and only people with extreme types of GD managed to visit our gender identity clinic for treatment. Currently, owing to media attention and the internet, it is easier to access information about our gender identity clinic, making the threshold lower to search for help. This could have led to referrals of people with milder forms of GD and people who were not sure of their feelings and just wanted to explore these with a psychologist. Such people eventually might not pursue HT. Another explanation might be that not all transgender people want to undergo HT, such as transmen or people with a non-binary identity who only want a mastectomy.[24]

By contrast, we noticed that the percentage of people who underwent gonadectomy within 5 years after the start of HT remained stable over time. At the start of the clinic in 1972, knowledge about transgender care was limited and only people who wished for a classic treatment, consisting of a diagnostic phase, HT plus social transitioning, and surgery (in this order), were treated. There was no room for partial treatments. Since the publication of the Standards of Care Version 6 in 2001, other types of treatment are offered.[25] In addition, in 2014, a change in Dutch law allowed transgender people without a wish to undergo gonadectomy to alter the sex on their birth certificate with a statement of an expert who declared that the individual was diagnosed with GD (Dutch civil law, article 1:28). Although these changes in clinical guidelines and the law might have led to a decrease in the number of transgender people choosing gonadectomy, the current results do not show this. However, the follow-up time of this study might be too short to notice such changes.

In the HT group, 22% of people who were eligible for surgery had not undergone gonadectomy. These numbers are

comparable with a study from Sweden[26] but larger than in a study from Belgium,[27] in which approximately 15% of transwomen and transmen did not undergo gonadectomy. A possible wish to carry a child could change these numbers in the future, because fertility has become a more important issue.

Despite the large increase in treated transgender people, the percentage of people who underwent gonadectomy but regretted their decision was still very small (0.5%). In a review by Pfäfflin[28] in 1992, regret was reported by less than 1% of transmen and 1% to 1.5% by transwomen after gonadectomy. More recent studies have reported regret percentages of 0%[29,30] to 2%[7] and 6%[31] after gonadectomy. 13 of the 14 people who regretted gonadectomy had started HT from 1978 through 1997 and 1 started in 2004. At best, this indicates that the diagnostic and eligibility criteria for treatment have improved over the past decade. Another explanation might be the altered treatment protocol, which also allowed people to receive HT without gonadectomy. Our findings could be an underestimation of people with regret after gonadectomy, because some might choose to go elsewhere for reversal therapy or might experience regret without pursuing reversal surgery or HT. Regret might not always result in a desire for reversal therapy, as it may be hidden from others. In addition, in our population the average time to regret was 130 months, so it might be too early to examine regret rates in people who started with HT in the past 10 years.

The Center of Expertise on Gender Dysphoria of the VUmc Amsterdam is the largest gender identity clinic in the Netherlands, where people of all ages, including children and adolescents, are treated. Life-time follow-up is recommended, making it a useful study population for collection of epidemiologic data and future long-term studies of treatment effects. However, there are some limitations. Because this is a retrospective chart review study, some data could be lacking. (i) Some people who once visited our clinic might not be reported in our database. However, we used several search strategies to identify the total study population, thereby decreasing the possibility of missing people. (ii) A large number of transgender people who had initially received treatment in our center were lost to follow-up. Although transgender people receive lifelong care, a large group (36%) did not return to our clinic after several years of treatment. Therefore, we could have missed some information on, for example, gonadectomies performed at other centers or people with regret.

CONCLUSIONS

We found that the prevalence of treated transgender people increased exponentially. Because of this growing population, it is necessary that health care providers outside university clinics also have knowledge about GD and its treatment, because HT can influence the course of several diseases[32,33] and interact with several types of medication.[34] We also found that of all transgender people treated with HT, approximately 22% kept their gonads in situ. These people require special attention, because the long-term effects of HT on the testes, ovaries, and uterus are not established. These topics and other possible complications, such as cancer risks, are subjects for further research.

ACKNOWLEDGMENT

The authors would like to thank Jos A.J. Megens for his role in accurate administrative work which helped us defining the study population.

Corresponding Author: M. den Heijer, MD, PhD, Department of Internal Medicine, Section Endocrinology, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, the Netherlands. Tel: +31-20-444-0530; Fax: +31-20-444-4313; E-mail: m.denheijer@vumc.nl

Conflicts of Interest: The authors report no conflicts of interest.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

STATEMENT OF AUTHORSHIP

Category 1

(a) Conception and Design
C.M. Wiepjes; N.M. Nota; C.J.M. de Blok; M. Klaver; R.T. de Jongh; P. Cohen-Kettenis; L.J.G. Gooren; B.P.C. Kreukels; M. den Heijer

(b) Acquisition of Data
C.M. Wiepjes; N.M. Nota; C.J.M. de Blok; M. Klaver; T.D. Steensma; B.P.C. Kreukels

(c) Analysis and Interpretation of Data
C.M. Wiepjes; N.M. Nota; C.J.M. de Blok; M. Klaver; A.L.C. de Vries; S.A. Wensing-Kruger; R.T. de Jongh; M. Bouman; T.D. Steensma; P. Cohen-Kettenis; L.J.G. Gooren; B.P.C. Kreukels; M. den Heijer

Category 2

(a) Drafting the Article
C.M. Wiepjes; N.M. Nota; P. Cohen-Kettenis; L.J.G. Gooren; B.P.C. Kreukels

(b) Revising It for Intellectual Content
N.M. Nota; C.J.M. de Blok; M. Klaver; A.L.C. de Vries; S.A. Wensing-Kruger; R.T. de Jongh; M. Bouman; T.D. Steensma; P. Cohen-Kettenis; L.J.G. Gooren; B.P.C. Kreukels; M. den Heijer

Category 3

(a) Final Approval of the Completed Article
C.M. Wiepjes; N.M. Nota; C.J.M. de Blok; M. Klaver; A.L.C. de Vries; S.A. Wensing-Kruger; R.T. de Jongh; M. Bouman; T.D. Steensma; P. Cohen-Kettenis; L.J.G. Gooren; B.P.C. Kreukels; M. den Heijer

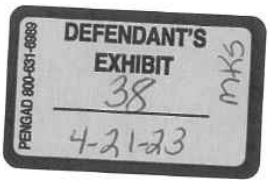
REFERENCES

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.

2. Wylie K, Knudson G, Khan SI, et al. Serving transgender people: clinical care considerations and service delivery models in transgender health. **Lancet 2016;388:401-411.**
3. Arcelus J, Bouman WP, Van Den Noortgate W, et al. Systematic review and meta-analysis of prevalence studies in transsexualism. **Eur Psychiatry 2015;30:807-815.**
4. De Cuypere G, Van Hemelrijck M, Michel A, et al. Prevalence and demography of transsexualism in Belgium. **Eur Psychiatry 2007;22:137-141.**
5. Esteva de Antonio I, Gomez-Gil E, Almaraz MC, et al. [Organization of healthcare for transsexual persons in the Spanish national health system]. **Gac Sanit 2012;26:203-209 [in Spanish].**
6. Caldarera A, Pfäfflin F. Transsexualism and sex reassignment surgery in Italy. **Int J Transgend 2011;13:26-36.**
7. Dhejne C, Oberg K, Arver S, et al. An analysis of all applications for sex reassignment surgery in Sweden, 1960–2010: prevalence, incidence, and regrets. **Arch Sex Behav 2014; 43:1535-1545.**
8. Bakker A, van Kesteren P, Gooren L, et al. The prevalence of transsexualism in the Netherlands. **Acta Psychiatr Scand 1993;87:237-238.**
9. Van Kesteren P, Asscheman H, Megens J, et al. Mortality and morbidity in transsexual subjects treated with cross-sex hormones. **Clin Endocrinol 1997;47:337-342.**
10. Cohen-Kettenis PT, Delemarre-van de Waal HA, Gooren LJ. The treatment of adolescent transsexuals: changing insights. **J Sex Med 2008;5:1892-1897.**
11. Kreukels BP, Cohen-Kettenis PT. Puberty suppression in gender identity disorder: the Amsterdam experience. **Nat Rev Endocrinol 2011;7:466-472.**
12. World Health Organization. The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. Geneva: World Health Organization; 1993.
13. Asscheman H, Giltay EJ, Megens JA, et al. A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones. **Eur J Endocrinol 2011; 164:635-642.**
14. Asscheman H, Gooren LJG, Eklund PLE. Mortality and morbidity in transsexual patients with cross-gender hormone treatment. **Metabolism 1989;38:869-873.**
15. Dekker MJ, Wierckx K, Van Caenegem E, et al. A European network for the investigation of gender incongruence: endocrine part. **J Sex Med 2016;13:994-999.**
16. Kreukels BP, Haraldsen IR, De Cuypere G, et al. A European network for the investigation of gender incongruence: the ENIGI initiative. **Eur Psychiatry 2012;27:445-450.**
17. Aitken M, Steensma TD, Blanchard R, et al. Evidence for an altered sex ratio in clinic-referred adolescents with gender dysphoria. **J Sex Med 2015;12:756-763.**
18. de Vries AL, Cohen-Kettenis PT. Clinical management of gender dysphoria in children and adolescents: the Dutch approach. **J Homosex 2012;59:301-320.**
19. Sullivan KM, Dean A, Soe MM. OpenEpi: a web-based epidemiologic and statistical calculator for public health. **Public Health Rep 2009;124:471-474.**
20. Kuyper L. Transgenders in Nederland: prevalentie en attitudes. **Tijdschr Seksuol 2012;36:129-135 [in Dutch].**
21. Kuyper L, Wijsen C. Gender identities and gender dysphoria in the Netherlands. **Arch Sex Behav 2014;43:377-385.**
22. Wallien MSC, Cohen-Kettenis PT. Psychosexual outcome of gender-dysphoric children. **J Am Acad Child Adolesc Psychiatry 2008;47:1413-1423.**
23. Ristori J, Steensma TD. Gender dysphoria in childhood. **Int Rev Psychiatry 2016;28:13-20.**
24. Beek TF, Kreukels BP, Cohen-Kettenis PT, et al. Partial treatment requests and underlying motives of applicants for gender affirming interventions. **J Sex Med 2015;12:2201-2205.**
25. De Cuypere G, Gijs L. Care for adults with gender dysphoria. In: Kreukels BPC, Steensma TD, de Vries ALC, eds. Gender dysphoria and disorders of sex development: progress in care and knowledge. New York: Springer Science + Business Media; 2014. p. 231-254.
26. Johansson A, Sundbom E, Hojerback T, et al. A five-year follow-up study of Swedish adults with gender identity disorder. **Arch Sex Behav 2010;39:1429-1437.**
27. Wierckx K, Elaut E, Declercq E, et al. Prevalence of cardiovascular disease and cancer during cross-sex hormone therapy in a large cohort of trans persons: a case-control study. **Eur J Endocrinol 2013;169:471-478.**
28. Pfäfflin F. Regrets after sex reassignment surgery. **J Psychol Hum Sex 1993;5:69-85.**
29. Vujovic S, Popovic S, Sbutega-Milosevic G, et al. Transsexualism in Serbia: a twenty-year follow-up study. **J Sex Med 2009;6:1018-1023.**
30. Lawrence AA. Factors associated with satisfaction or regret following male-to-female sex reassignment surgery. **Arch Sex Behav 2003;32:299-315.**
31. Imbimbo C, Verze P, Palmieri A, et al. A report from a single institute's 14-year experience in treatment of male-to-female transsexuals. **J Sex Med 2009;6:2736-2745.**
32. Johnson EL, Kaplan PW. Caring for transgender patients with epilepsy. **Epilepsia 2017;58:1667-1672.**
33. Tauboll E, Sveberg L, Svalheim S. Interactions between hormones and epilepsy. **Seizure 2015;28:3-11.**
34. Polderman KH, Gooren LJG, Asscheman H, et al. Induction of insulin resistance by androgens and estrogens. **J Clin Endocrinol Metab 1994;79:265-271.**

frontiers in Psychiatry

ORIGINAL RESEARCH
published: 29 March 2021
doi: 10.3389/fpsyt.2021.632784



A Follow-Up Study of Boys With Gender Identity Disorder

Devita Singh[1], *Susan J. Bradley*[2] *and Kenneth J. Zucker*[2]*

[1] Department of Human Development and Applied Psychology, Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, Canada. [2] Department of Psychiatry, University of Toronto, Toronto, ON, Canada

This study reports follow-up data on the largest sample to date of boys clinic-referred for gender dysphoria ($n = 139$) with regard to gender identity and sexual orientation. In childhood, the boys were assessed at a mean age of 7.49 years (range, 3.33–12.99) at a mean year of 1989 and followed-up at a mean age of 20.58 years (range, 13.07–39.15) at a mean year of 2002. In childhood, 88 (63.3%) of the boys met the DSM-III, III-R, or IV criteria for gender identity disorder; the remaining 51 (36.7%) boys were subthreshold for the criteria. At follow-up, gender identity/dysphoria was assessed via multiple methods and the participants were classified as either persisters or desisters. Sexual orientation was ascertained for both fantasy and behavior and then dichotomized as either biphilic/androphilic or gynephilic. Of the 139 participants, 17 (12.2%) were classified as persisters and the remaining 122 (87.8%) were classified as desisters. Data on sexual orientation in fantasy were available for 129 participants: 82 (63.6%) were classified as biphilic/androphilic, 43 (33.3%) were classified as gynephilic, and 4 (3.1%) reported no sexual fantasies. For sexual orientation in behavior, data were available for 108 participants: 51 (47.2%) were classified as biphilic/androphilic, 29 (26.9%) were classified as gynephilic, and 28 (25.9%) reported no sexual behaviors. Multinomial logistic regression examined predictors of outcome for the biphilic/androphilic persisters and the gynephilic desisters, with the biphilic/androphilic desisters as the reference group. Compared to the reference group, the biphilic/androphilic persisters tended to be older at the time of the assessment in childhood, were from a lower social class background, and, on a dimensional composite of sex-typed behavior in childhood were more gender-variant. The biphilic/androphilic desisters were more gender-variant compared to the gynephilic desisters. Boys clinic-referred for gender identity concerns in childhood had a high rate of desistance and a high rate of a biphilic/androphilic sexual orientation. The implications of the data for current models of care for the treatment of gender dysphoria in children are discussed.

Keywords: gender dysphoria, gender identity disorder, gender non-conformity, sexual orientation, DSM-5

OPEN ACCESS

Edited by:
Maria Inês Rodrigues Lobato, Clinical Hospital of Porto Alegre, Brazil

Reviewed by:
Ray Blanchard, University of Toronto, Canada
Lanna Petterson, University of Lethbridge, Canada
Scott William Semenyna, University of Lethbridge, Canada

***Correspondence:**
Kenneth J. Zucker
ken.zucker@utoronto.ca

Specialty section:
This article was submitted to Public Mental Health, a section of the journal Frontiers in Psychiatry

Received: *24 November 2020*
Accepted: *18 February 2021*
Published: *29 March 2021*

Citation:
Singh D, Bradley SJ and Zucker KJ (2021) A Follow-Up Study of Boys With Gender Identity Disorder. Front. Psychiatry 12:632784. doi: 10.3389/fpsyt.2021.632784

INTRODUCTION

Gender identity is considered to be, for most people, a central aspect of one's sense of self (1–6).[1] By around 3 years of age, if not earlier, most children can self-label themselves as either a boy or a girl (11–14) although cognitive-developmental gender theory suggests that the understanding of gender as an "invariant" aspect of the self does not occur until early to middle childhood, with the achievement of concreate operational thought (12, 15, 16). Gender differences in the adoption of gender role behavior, i.e., behavior associated with cultural definitions of masculinity and femininity, also emerge during the preschool years, if not earlier. These behaviors span various domains, including peer, toy, role play, and activity preferences [e.g., (3, 17, 18)]. Normative developmental research has long documented that, on average, both gender identity and gender role behaviors show significant and substantial between-sex differences (19–21). Later in development, sexual orientation also shows a substantial between-sex difference, i.e., most males are sexually attracted to females and most females are sexually attracted to males (19, 22).

In the 1950s and 1960s, a small clinical literature began to describe the phenomenology of children who displayed marked gender-variant behavior, including the strong desire to be of the other gender [e.g., (23–27)]. Subsequent volumes by Stoller (28) and Green (29) provided more comprehensive descriptions of such children. These early works were the sequel to the introduction of the diagnostic term Gender Identity Disorder (GID) of Childhood to the psychiatric nomenclature in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* [DSM-III; (30)], currently termed Gender Dysphoria (GD) in the DSM-5 (31). Since 1980, empirical research has examined a number of parameters pertaining to GID/GD: epidemiology, diagnostic and assessment methods, associated psychopathology, causal mechanisms, and therapeutic approaches [for reviews, see, e.g., (32–39)].

An additional parameter (the focus of the present study) pertains to the developmental course of GID in children. In the early literature, it was posited by some that pervasive gender-variant behavior in children might be a predictor of GID in adulthood (termed Transsexualism in the DSM-III) [e.g., (26, 40)]. At the same time, it was also recognized that gender-variant behavior in childhood was associated with sexual orientation (in males, androphilia, i.e., sexual attraction to men; in females, gynephilia, i.e., sexual attraction to women), but without co-occurring gender dysphoria [see, e.g., (41, 42); for a meta-analytic review, see (43)].

To date, there have been at least 10 follow-up studies of children whose behavior was consistent with the DSM diagnosis of GID (or GD per DSM-5) (44–53). Across these studies, the year at the time of first evaluation in childhood ranged from 1952 (49) to 2008 (51). For the 9 studies that included boys, the sample sizes (excluding those lost to follow-up) ranged from 6 to 79 (Mean age, 26 years). Most of these studies also provided the age at the time of first evaluation in childhood, which ranged from a mean of 7 years (47) to a mean of 9 years (48), with an age range from 4 to 12 years.

At the time of follow-up, using different metrics (e.g., clinical interview, maternal report, dimensional measurement of gender dysphoria, a DSM diagnosis of GID, etc.), these studies provided information on the percentage of boys who continued to have gender dysphoria (herein termed "persisters") and the percentage of boys who did not (herein termed "desisters").[2] Of the 53 boys culled from the relatively small sample size studies (Bakwin, Davenport, Kosky, Lebovitz, Money and Russo, Zuger), the percentage classified as persisters was 9.4% (age range at follow-up, 13–30 years). In Green (47), the percentage of persisters was 2% (total $n = 44$; Mean age at follow-up, 19 years; range, 14–24); in Wallien and Cohen-Kettenis (52), the percentage of persisters was 20.3% (total $n = 59$; Mean age at follow-up, 19.4 years; range, 16–28); and in Steensma et al. (51), the percentage of persisters was 29.1% (total $n = 79$; Mean age at follow-up, 16.1 years; range, 15–19). Across all studies, the percentage of persisters was 17.4% (total $N = 235$), with a range from 0 to 29.1%.[3]

These studies also provided information on the sexual orientation of the boys at the time of follow-up. In the early studies, sexual orientation was ascertained from various sources (e.g., open-ended interviews with the patient, parent-report, chart information, etc.). In the more recent studies, sexual orientation was assessed in a more systematic manner, such as the use of a structured interview to assign a Kinsey-based rating of sexual orientation in fantasy and a rating of sexual orientation in behavior, dummy coded where a 0 = gynephilia and a 6 = androphilia [e.g., (47)].

Of the 53 boys culled from the relatively small sample size studies (op. cit.), 13 (34.2%) of the patients were classified as gynephilic and 25 (65.8%) were classified as biphilic/androphilic.[4] In the remaining 15 patients (28.3% of the combined samples), their sexual orientation was either uncertain or unknown.

[1] In one study, Turner and Brown (7) found that school-age children rarely mentioned their gender when providing open-ended self-descriptions; the most frequent descriptor pertained to activities and preferences. Turner and Brown suggested that it might be the case that gender is so central to one's self-concept that it "goes without saying" (p. 709). In contemporary times in the West, a very small number of parents choose to not "gender" their children ("theybies") by not referring to them as boys or girls (and, at times, not even announcing to others the child's biological sex), dressing them in gender-neutral ways, etc. Little is known about the gender identity and gender role patterns of these children (8–10).

[2] The terms persistence and desistance have been used for a long time in clinical developmental psychiatry and psychology [e.g., (54)]. Zucker (55) was the first to apply these terms to describe the developmental psychosexual trajectories of children diagnosed with GID.

[3] The percentages provided here differ somewhat from other summary reviews [(39), pp. 285–286, (56, 57)] because we have excluded patients who were seen for the first time in adolescence [for this reason, data from Zuger (58) are also not included]. One other follow-up study was conducted by Nakamura (59). Unfortunately, this dissertation is not available for purchase at ProQuest (Ann Arbor, MI) and is only available for loan at the University of Essex library. Due to COVID-19 restrictions, it is currently inaccessible (K. Clarke, personal communication to G. Rieger, June 15, 2020). The director of the clinic at the time when the data were collected does not have a copy of the dissertation (D. Di Ceglie, personal communication, June 15, 2020).

[4] As pointed out by Reviewer 1, biphilic is a dubious neologism, combining Latin and Greek derivatives. Diphilic would be the more accurate derivative. However, introducing this term would probably confuse many readers, so we have retained the term biphilic (see https://en.wikipedia.org/wiki/Androphilia_and_gynephilia).

In Green's (47) study, 11 (25%) of the boys were classified as gynephilic (Kinsey ratings of 0–1) and 33 (75%) were classified as biphilic/androphilic in fantasy (Kinsey ratings of 2–6). For behavior, 6 (20%) were classified as gynephilic and 24 (80.0%) were classified as biphilic/androphilic. The remaining 14 boys (31.8% of the total sample) could not be classified with regard to behavior because they had had no interpersonal sexual experiences. In Green's study, the sexual orientation of a comparison group of boys, who had been recruited from the community, was also assessed: 100% of these boys ($n = 35$) were classified as gynephilic in fantasy and 96% ($n = 25$) were classified as gynephilic in behavior.

In the Wallien and Cohen-Kettenis (52) study, sexual orientation was assessed for attraction (2 items), fantasy (2 items), behavior (4 items), and sexual identity (1 item) using a self-developed Sexual Orientation Questionnaire. As in Green, Kinsey-type ratings were used in the analysis. Depending on the metric, data on sexual orientation were not available for anywhere between 22 and 40 (27.2–67.7%) patients. For attraction, 32% were classified as gynephilic and 68% were classified as androphilic (total $N = 37$); for fantasy, 19% were classified as gynephilic, 19% were classified as biphilic, and 62% were classified as androphilic (total $N = 21$); for behavior, 21% were classified as gynephilic, 16% were classified as biphilic, and 63% were classified as androphilic (total $N = 19$); lastly, for sexual identity, 19% were classified as gynephilic ("heterosexual"), 19% were classified as biphilic ("bisexual"), and 62% were classified as androphilic ("homosexual") (total $N = 27$). Steensma et al. (51) used the same metrics as Wallien and Cohen-Kettenis. Depending on the metric, data on sexual orientation were not available for anywhere between 25 and 40 (31.6%-50.6%) patients. For attraction, 19.2% were classified as gynephilic, 15.4% were classified as biphilic, and 65.4% were classified as androphilic (total $N = 52$); for fantasy, 14% were classified as gynephilic, 22% were classified as biphilic, and 64% were classified as androphilic (total $N = 50$); for behavior, 35.9% were classified as gynephilic, 12.8 were classified as biphilic, and 51.3% were classified as androphilic (total $N = 39$); lastly, for sexual identity, 13% were classified as gynephilic ("heterosexual"), 27.8% were classified as biphilic ("bisexual"), and 59.3% were classified as androphilic ("homosexual") (total $N = 54$).

In recent years, there have been various criticisms of these follow-up studies [see, e.g., (60–63); for a rebuttal, see (64)], particularly with regard to the putatively high percentage of desistance. It has been questioned, for example, to what extent the patients in these studies truly had GID/GD. For example, in the early studies, prior to the publication of DSM-III, one could reasonably argue that the diagnostic status of the patients was unclear because there were no formal diagnostic criteria to rely upon. However, one could argue in return that the behavior of these boys was phenomenologically consistent with the subsequent DSM criteria.

Consider, for example, the systematic study by Green [(47), Figure 1.2]. Green reported that 15% of the feminine boys, per parent-report, had "never" expressed the desire to be a girl or a woman at the time of the baseline assessment, 60% "occasionally" had such a desire, and only 25% had such a desire "frequently." Thus, a conservative critic might argue that only the last group would have met one of the key indicators for the GID/GD diagnosis in the DSM.[5] On the other hand, suppose a boy "occasionally" voiced the desire to be a girl over a period of several years. One might want to make the case that this would be consistent with the DSM descriptors of "persistently" or "repeatedly," etc. Of course, one could debate what would genuinely count as "occasionally" (in Green's trichotomous metric, it would be anything more than "never" and less than "frequently"). In any case, it is probably reasonable to argue that, in Green's study, some boys were threshold and some boys were subthreshold for the equivalent of a DSM diagnosis. Given that in Green's study only one boy persisted with gender dysphoria at the time of follow-up, the threshold-subthreshold distinction would not really matter.

Studies that employed DSM criteria for GID/GD allow for a more formal examination of the "No True Scotsman" argument (https://en.wikipedia.org/wiki/No_true_Scotsman).

In the Wallien and Cohen-Kettenis (52) study, the DSM-III-R criteria were used to diagnose GID. Of the 12 persisters, all met the criteria for GID at the time of the baseline assessment; in contrast, only 68% of the 47 desisters met the criteria for GID; the remainder were deemed subthreshold for the diagnosis. Thus, in their study, the threshold-subthreshold distinction appears to have been an important one in predicting outcome; nonetheless, it should be noted that 68% of the desisters had been threshold for the diagnosis in childhood—perhaps a strong rebuttal to the No True Scotsman argument. In Steensma et al. (51), the DSM-IV-TR criteria were used. Of the 23 persisters, 21 (91.3%) met the criteria for GID; in contrast, only 22 (39.3%) of the 56 desisters were threshold for the diagnosis, suggesting an even more substantial difference in the threshold-subthreshold distinction than was found in Wallien and Cohen-Kettenis. Although the latter percentage was lower than what was found in Wallien and Cohen-Kettenis, that almost 40% of the desisters met the criteria for GID in childhood still argues in favor that the children were desisting from something.[6]

From Wallien and Cohen-Kettenis (52) and Steensma et al. (51), one predictor of outcome, therefore, was the distinction between being threshold or subthreshold for the GID diagnosis in childhood. Dimensional measures of gender-variant behavior have also proven useful. In both Wallien and Cohen-Kettenis and Steensma et al., dimensional measures of sex-typed behavior in childhood also significantly discriminated between the persisters and desisters, with the former group having, on average, more severe gender-variant behavior at the time of the childhood

[5]The situation is compounded even further because in the DSM-IV, unlike in the DSM-III and DSM-III-R (65), the stated desire to be of the other gender was not a necessary criterion for the diagnosis [for the rationale, see (66), pp. 483–486]. In DSM-5, the desire to be of the other gender does not require explicit verbalization: the clinician is allowed leeway in drawing inferences based on other sources of information [see (67), pp. 904–905].

[6]In the follow-up study by Drummond et al. (46) of 25 girls from our clinic, the desistance rate was 88%. Of the 22 desisters, 13 (59.0%) met the DSM-III, III-R or IV criteria for GID. In Wallien and Cohen-Kettenis (52), of the 9 girls who desisted, 55.5% met the DSM-III-R criteria for GID. In Steensma et al. (51), of the 24 girls who desisted, 58.3% met the DSM-IV criteria for GID.

assessment. Steensma et al. found two other predictors of persistence: boys who were assessed at an older age and boys who had made either a partial or complete gender "social transition" [see (68–70)]. Of the 12 boys who had partially or completely transitioned prior to puberty, 10 (83.3%) were classified as persisters. In contrast, of the 67 boys who had not socially transitioned, only 13 (19.4%) were classified as persisters.

In the present study, we provide follow-up data with regard to both gender identity (persistence vs. desistance) and sexual orientation (gynephilia vs. biphilia/androphilia) on the largest sample of boys studied to date. Apart from providing percentage data on these two variables, which will be discussed in a comparative perspective in relation to the prior studies and the epidemiological literature, we also examine the predictors of outcome in relation to both demographic and sex-typed behavior measures (including whether or not the boys were threshold or subthreshold for GID) collected at the time of the baseline assessment in childhood.

METHOD

Participants

The participants were 139 boys ("birth-assigned males")[7] who, in childhood, had been referred to and then assessed in the Gender Identity Service, Child, Youth, and Family Program at the Centre for Addiction and Mental Health (CAMH) in Toronto, Ontario between 1975 and 2009 (Mean year of assessment, 1989.36) and were adolescents or adults at follow-up (Mean year at follow-up, 2002.35).[8]

Participants entered the follow-up study through two methods of recruitment. The majority of participants (77%) were recruited for research follow-up. There were two main waves of participant recruitment through research contact, from 1986 to 1993 (n = 32) and then from 2009 to 2011 (n = 71). During the period of data collection, 32 patients re-contacted the service for clinical reasons (eight for gender dysphoria, six for sexual orientation, and 18 for heterogeneous concerns) [for details, see (77), Appendix E]. They were informed about the opportunity to participate in the follow-up study and subsequently completed the study protocol. The majority of the patient-initiated participants had contacted the clinic between the two main waves of research recruitment. Thus, from 1994 to 2008, the participants who entered the study were primarily those who had contacted the service for clinical reasons.

In the early wave of follow-up, a lower-bound age for participation was set at 14 years, but by the mid-1990s this was changed to a lower-bound age of 16 years. In total, 110 (79.1%) participants were at least 16 years of age and 29 (20.9%) were younger than 16. Across the entire period of data collection, eligible participants, after review of the medical chart, were contacted at random (other than the participants who had returned to the service for clinical reasons). Due to lack of study resources and time constraints, contact with 162 other eligible participants was not attempted.

In total, 145 patients were approached about the follow-up study, either through research contact (n = 113) or following their clinical involvement with the Gender Identity Service (n = 32). Six patients declined, which yielded a participation rate of 95.9%. For those recruited for research purposes, initial contact, by telephone, letter or email, was first made with the parents because the patients were minors at the time of the childhood assessment and may have had no recollection of their clinic attendance. A total of 19 (14.3%) potential participants could not be reached/traced through previous addresses, registrars, and personal contacts.

Of the 139 participants, 110 were seen for a face-to-face assessment. For various reasons, the remaining 29 patients could not be seen for the face-to-face assessment (e.g., lived in another province or country, "too busy," severe mental health issues). For some patients, they provided some information over the phone or information was provided by the parents; thus, for these patients, it was possible to obtain some follow-up data about their gender identity and sexual orientation.

The demographic characteristics of the participants, including their age at assessment in childhood and at the time of follow-up, are shown in **Table 1**. The GID diagnosis in childhood was based on the DSM-III (n = 53), DSM-III-R (n = 46), or DSM-IV (n = 40) criteria applicable at the time of assessment.[9] A total of 88 (63.3%) boys met complete DSM criteria for GID in childhood. The remaining 51 (36.7%) boys were subthreshold for a DSM diagnosis, but all had some indicators of GID, and, based on the historical information provided during the assessment, some would have met the complete DSM criteria at some point in their lives prior to their assessment in childhood.[10] The percentage who met the complete DSM criteria for GID did not differ significantly as a function of DSM edition, $\chi^2_{(2)} < 1$.

Procedure

The majority of participants who completed the face-to-face assessment were evaluated on a single day. Three participants were seen twice. In these instances, the participants completed the self-report measures during their second visit as the complexity of their clinical presentation extended the duration of the assessment. Participants were provided a stipend for their participation in the follow-up assessment and reimbursement for travel expenses. For participants followed-up prior to 2009 (n = 68), the data were collected by the third author; for those followed-up between 2009 and 2011, the data were collected

[7]Two reviewers asked why we chose to use the noun "boys" instead of the noun "males." In our view, the question was reasonable but also a matter of semantics and taste. The third edition of *The Oxford Dictionary of Current English* (71) defines boy as "a male child..." Thus, we believe that the two words can be used synonymously. Males can refer to any age in the life-span whereas boys connote childhood. The participants in our study were coded as male at the time of their birth in the hospital delivery record, of which we had the actual birth records for the majority of the participants in the current study (72). As per Bouman et al. (73), one would say that the participants were "assigned male at birth" and then declared socially to be "boys" (74).

[8]The clinic was established in 1975 at the Clarke Institute of Psychiatry (75, 76), which became part of the CAMH in 1998.

[9]For boys seen prior to the publication of DSM-III in 1980, the draft criteria were used.

[10]In DSM-III, termed Atypical Gender Identity Disorder; in DSM-III-R and DSM-IV, termed Gender Identity Disorder Not Otherwise Specified.

TABLE 1 | Demographic characteristics ($N = 139$).

Characteristic	*M*	*SD*	Range	%
From childhood				
Age (in years)	7.49	2.66	3.33–12.99	
Year of birth	1981.87	7.50	1966–1996	
Year of assessment	1989.36	7.50	1975–2004	
IQ[a]	105.93	15.47	69–138	
Social class[b]	40.74	15.15	8.0–66.0	
Marital status[c]				
Two-parent family				64.7
Other				35.3
Caucasian				84.9
At follow-up				
Age (in years)	20.58	5.22	13.07–39.15	
Year of follow-up	2002.35	9.08	1986–2011	
Follow-up interval (in years)[d]	12.86	6.07	2.77–29.29	
IQ[e,f]	105.88	16.03	65–138	

[a] *Full-Scale IQ was obtained with age-appropriate Wechsler intelligence scales.*
[b] *Hollingshead's (79) Four Factor Index of Social Status (absolute range, 8–66).*
[c] *Other included the following family constellations: single parent, separated, divorced, living with relatives, or in the care of a child protection agency.*
[d] *Interval denotes the time between childhood assessment and follow-up assessment.*
[e] *Full Scale IQ estimated using four subtests: Vocabulary, Comprehension, Block Design, and Object Assembly.*
[f] *An IQ score was available only for participants who completed the face-to-face assessment. Of these, scores were not available for one participant.*

by the first author ($n = 71$). The study was approved by the Institutional Review Boards at the Clarke Institute of Psychiatry (subsequently the Centre for Addiction and Mental Health; Protocol #198/2008–2011) and the University of Toronto.

Measures

Below, we describe the measures from assessment and follow-up of relevance for this article. A list of all measures used in the follow-up study can be found in Singh [(77), Table 4].

Childhood Assessment

Cognitive Functioning

Based on the child's age at the time of assessment, the appropriate version of the Wechsler Intelligence Scale for Children was administered (WPPSI-R or the WISC-R/WISC-III/WISC-IV). Full scale IQ scores were used to characterize level of cognitive functioning.

Behavioral and Emotional Problems

Parents completed the Child Behavior Checklist (CBCL), a measure of behavioral and emotional problems (78). Although not the focus of the present study, it is noted here because we used three CBCL indices (sum of all behavior problems and Internalizing and Externalizing *T* scores) as part of an internal validity analysis when comparing participants vs. non-participants (see Results).

Sex-Typed Behavior

Five child informant and two parent informant measures were used to assess the participants' sex-typed behavior in childhood: (1) Draw-a-Person [DAP] test (80); (2) a free-play task (81); (3) the Playmate and Playstyle Preferences Structured Interview (PPPSI) (82, 83); (4) sex-typed responses on the Rorschach test (84); (5) the Gender Identity Interview for Children (GIIC) (85–87); (6) the Gender Identity Questionnaire for Children (GIQC) (88–90); and (7) a measure of activity level/extraversion [(39); see also (91)]. These child and parent informant measures all have established discriminant validity, that is, they significantly differentiated the boys clinic-referred for gender identity concerns from control boys [for reviews, see (18, 92)]. A Childhood Sex-Typed Behavior Composite was subsequently computed for each participant (see below).

Follow-Up Assessment

Cognitive Functioning

Four subtests from the age-appropriate version of the Wechsler Intelligence Scales were administered (Vocabulary, Comprehension, Block Design, and Object Assembly). The standard scores from the subtests were averaged to form a prorated IQ score for cognitive functioning (93).

Concurrent Gender Identity

Concurrent gender identity was evaluated using a semi-structured interview and self-report questionnaires. During an audiotaped interview, each participant was asked to describe their current feelings about being a biological male. They were also asked to describe positive and negative aspects about their gender identity. For example, participants who reported a "male" gender identity were asked to describe positive and negative aspects of being male. The semi-structured interview also included questions based on the adolescent and adult GID criteria outlined in the DSM-III-R or DSM-IV (65, 94). Participants were asked to respond to these questions according to the last 12 months with *No*, *Sometimes*, or *Yes* [for details, see (77), Appendix G].

Two self-report measures were also used to assess current gender identity and gender dysphoria: (1) The Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (GIDYQ-AA) (95–97) or (2) the Gender Dysphoria/Identification questionnaire (GDIQ) (98). The GDIQ was developed prior to the GIDYQ-AA. As such, the GIDYQ-AA was introduced to the protocol subsequent to the GDIQ and, as a result, the more recent participants completed the GIDYQ-AA while earlier participants completed the GDIQ.

The male version of the GIDYQ-AA was completed. This 27-item questionnaire measures gender identity and gender dysphoria in adolescents or adults; participants over the age of 17 completed the adult version and younger participants completed the adolescent version. The adolescent and adult versions are identical except that, in the adult version, the words "man" and "woman" are used instead of "boy" and "girl." Each item was rated on a 1–5 point response scale with verbal anchor points ranging from *Never* to *Always* based on a time frame of the past 12 months. Coding was such that a "lower" score signified more gender dysphoria. Item examples include the following:

"In the past 12 months, have you felt unhappy about being a man?" and "In the past 12 months, have you had the wish or desire to be a woman?" Principal axis factor analysis identified a one-factor solution that accounted for 61.3% of the variance. All factor loadings were ≥0.30 (median, 0.86; range, 0.34–0.96). The GIDYQ-AA has strong evidence for discriminant validity and a high threshold for specificity (i.e., low false positive rate for non-GID individuals) [see (95, 96, 99–102)].

The GDIQ (98) contains 8 items pertaining to gender identity and gender dysphoria. Factor analysis identified two factors, accounting for 31.4 and 12.5% of the variance, respectively (all factor loadings ≥0.45). Factor 1 consisted of five items pertaining to gender dysphoria and Factor 2 consisted of three items pertaining to gender role identification. For the present study, only the questions for Factor 1 were used. Each item was rated on a 3-point or 5-point scale for the past 12 months (see **Appendix 1** in **Supplementary Material**).

Participants were classified as having persistent gender dysphoria if their mean score on the GIDYQ-AA was ≤3.00, in line with sensitivity and specificity analyses from other data sets (95, 96). For participants who did not complete the GIDYQ-AA, the GDIQ was used. A participant was classified as a persister if two or more of the following five items on the GDIQ were endorsed: wish to have been born a girl (Item 1), wish to have surgery to change body (Item 2), feel more like a girl than a boy (Item 3), wonder if would be happier as a girl (Item 4), and somewhat or very dissatisfied with being a boy (Item 5).

Information regarding participants' gender identity/gender dysphoria was also obtained during the semi-structured clinical interview and, therefore, allowed for cross-validation of these questionnaire data. For those participants who did not complete the face-to-face interview, clinical information regarding gender identity/gender dysphoria was obtained through self- or parent-report or chart review. Across the entire sample, the GIDYQ-AA was used to classify persistence or desistence for 64 participants, the GDIQ for 42 participants, and interview/chart data/parent report for 33 cases.

Sexual Orientation

Sexual orientation in fantasy was assessed with specific questions from an audiotaped face-to-face interview and the self-report Erotic Response and Orientation Scale (EROS) (103).

The interview asked about four types of sexual fantasy over the past 12 months: (1) crushes on other people; (2) sexual arousal to visual stimuli (e.g., acquaintances, partners, and individuals from movies, television, etc.); (3) sexual content of night dreams; and (4) sexual content of masturbation fantasies. During the interview, participants were not asked directly about the gender of the person or persons who elicited sexual arousal, thus allowing time for the participant to provide this information spontaneously. Directed questions about the gender of the person(s) who elicited sexual arousal were asked only if the participant did not volunteer specific information about whether their arousal was directed to same-sex or opposite-sex individuals, or both. By the end of the interview, each participant provided information about sexual arousal to both same-sex and opposite-sex individuals. Using the Kinsey scale criteria (104), the interviewer assigned Kinsey ratings that ranged from 0 (exclusively gynephilic in fantasy) to 6 (exclusively androphilic in fantasy) for each question. A dummy score of 7 denoted that the participant did not experience or report any fantasies. A global fantasy score was also derived based on ratings from the four questions. Kinsey ratings for sexual orientation in fantasy were available for 129 participants.

Inter-rater reliability on Kinsey ratings for sexual orientation in fantasy was examined for 29 participants, selected at random. The second scorer listened to the audio recordings of the semi-structured interview, with specific attention to the information collected on sexual orientation. The inter-rater agreement on the Kinsey global fantasy rating was very good (kappa = 0.95) and the kappa values for the four specific components ranged from 0.81 to 1.00.

The EROS is a 16-item self-report measure assessing sexual orientation in fantasy over the past 12 months. Half of the questions pertained to gynephilic fantasy (e.g., "How often have you noticed that you had sexual feelings [even the slightest] while looking at a woman?") and the other half pertained to androphilic fantasy (e.g., "How often have you noticed that you had sexual feelings [even the slightest] while looking at a man?"). Participants who were 18 years and older completed the adult version and younger participants completed the adolescent version. The adolescent and adult versions are identical except that, in the adult version, the words "man" and "woman" were used instead of "boy" and "girl." Each item was rated on a 5-point scale for frequency of occurrence, ranging from 1 ("none") to 5 ("almost every day"). Mean androphilic and gynephilic fantasy scores were derived for each participant. In the present study, we calculated a difference score between the participants' mean androphilic and gynephilic scores. Previous use of the EROS has shown good evidence of discriminant validity (98, 101).

Sexual orientation in behavior was assessed with specific questions during the face-to-face interview and with a modified version of the Sexual History Questionnaire (SHQ) (105). In the interview, questions asked about five types of sexual behavior: (1) dating; (2) holding hands in a romantic manner; (3) kissing; (4) genital fondling or touching a woman on the breasts, and (5) intercourse (penile-vaginal and anal). Kinsey ratings for behavior in the past 12 months were made in the same manner as fantasy ratings. Kinsey ratings for sexual orientation in behavior were available for 108 participants. Inter-rater reliability on Kinsey ratings for sexual orientation in behavior was examined for the same 29 participants. There was perfect inter-rater agreement on the Kinsey global behavior rating (kappa = 1.0) and the kappa values for the five specific components ranged from 0.91 to 1.00.

The modified SHQ consists of 20 questions. Ten questions pertained to gynephilic experiences (e.g., "How many women have you kissed on the lips in a romantic way?") and 10 questions pertained to androphilic experiences (e.g., "How many men have you kissed on the lips in a romantic way?"). Participants who were 18 years and older completed the adult version and younger participants completed the adolescent version. The adolescent and adult versions are identical except that, in the adult version, the words "man" and "woman" were used instead of "boy" and "girl." Each item was rated on a 5-point scale for frequency

of occurrence, ranging from 1 ("none") to 5 ("11 or more"), based on a time frame of the past 12 months. Mean total scores for gynephilic and androphilic experiences were derived. In the present study, we calculated a difference score between the participants' mean androphilic and gynephilic scores.

On the basis of Kinsey ratings, participants who completed the face-to-face interview were classified, similar to Green (47), into the following three sexual orientation groups for both fantasy and behavior: (1) gynephilic (Kinsey global ratings of 0–1); (2) biphilic/androphilic (Kinsey global ratings of 2–6), and (3) no sexual fantasy or behavior.

Social Desirability

Social desirability refers to the desire to cast a favorable impression on others. It can threaten the validity of self-report scales if in answering questions respondents seek social approval or try to represent themselves in a favorable manner (106). People scoring high on social desirability tend to provide socially acceptable answers regardless if their response accurately describes them. Participants 18 years and older completed the Marlow-Crowne Social Desirability Scale (M-CSDS) (107), which consists of 33 true-false items. The scale contains 18 culturally acceptable but unlikely statements keyed in the true direction and 15 socially undesirable but probable statements keyed in the false direction for a maximum possible score of 33. Participants 17 years and under were given a shorter version of the M-CSDS (108), containing 20 items that consist of 12 culturally acceptable but improbable statements keyed in the true direction and eight socially undesirable but probable statements keyed in the false direction for a maximum possible score of 20. For the present study, the percentage of endorsed socially desirable items was calculated for each participant. In order to integrate the data from both versions of the M-CSDS, participants' percentage score on each measure was converted to a proportion score which ranged from 0 to 1, which was used in all analyses. A higher proportion score indicates a greater propensity to give socially desirable responses. Several studies have found that the MCSDS is a reliable and valid measure of social desirability (107, 109, 110).

RESULTS

Preliminary Analyses

Participants vs. Non-participants

Given that not all eligible participants were seen for follow-up, it is important to see to what extent the participants vs. non-participants were similar with regard to baseline characteristics, in part to gauge the internal validity of the sample (111).

The non-participants consisted of three subgroups: (1) patients who were eligible to participate in the study but were not contacted ($n = 163$), (2) patients who declined to participate ($n = 6$), and (3) patients who were not successfully traced ($n = 19$). Two sets of analyses were conducted to compare study participants vs. non-participants. First, the participants were compared to the patients who were eligible but not contacted. Second, the participants were compared to those who declined to participate and to those where contact was attempted but not successfully traced. Group comparisons were conducted on five demographic variables (age at assessment in childhood, IQ, ethnicity, and parents' marital status and social class), parent-report of behavior problems on the CBCL (three indices), and nine measures of childhood sex-typed behavior.

Of these 17 variables, there was only one significant difference between the 139 boys in the study compared to the 163 boys who were eligible to participate but were not contacted: participants had a higher IQ than non-participants, $t_{(289)} = 2.01$, $p = 0.046$.[11] The effect size for this comparison was small (unpooled $d = 0.22$) [for details, see (77), Tables 5, 6]. When compared to the six cases where participation in the study was declined and to the 19 cases where the families could not be traced, there was also only one significant difference: parent's marital status, $\chi^2_{(2)} = 9.02$, $p = 0.011$. The participants did not differ significantly from the non-participants who refused; however, they differed significantly from the cases that could not be traced, $\chi^2_{(1)} = 6.39$, $p = 0.012$. The participants were more likely to have originated within a two-parent household than those who could not be traced. The comparison between the non-participants who refused and those who could not be traced approached significance ($p = 0.056$, Fisher's exact test). Again, the non-participants who could not be traced were more likely to have come from a family composition that was not two-parent. A further summary of comparisons between the participants and those who declined or could not be traced can be found in the **Supplementary Material**.

Participants: Method of Recruitment

Using *t*-tests or chi-square tests, the 107 participants who entered the study through research contact were compared to the 32 participants who were recruited into the study after they had re-contacted the clinic for clinical reasons on the demographic variables, CBCL behavior problems in childhood, and the measures of childhood sex-typed behavior. There were no significant differences between the two groups on the demographic variables of age at assessment, ethnicity or parents' social class and marital status (*p*s > 0.05). The comparison on childhood IQ approached significance, $t_{(137)} = 1.97$, $p = 0.051$, with the research entry participants having, on average, a higher IQ than the clinical entry participants. On the CBCL, there was a significant difference on Internalizing problems only, $t_{(137)} = -2.02$, $p = 0.046$, with the clinical entry participants rated by their parents as having more internalizing problems compared to the research entry participants. Of the nine measures of childhood sex-typed behavior, the two groups differed significantly on three: (1) free play, $t_{(119)} = -2.11$, $p = 0.037$, (2) the Gender Identity Interview for Children, $t_{(83)} = -2.09$, $p = 0.04$, and (3) the Gender Identity Questionnaire for Children, $t_{(95)} = 2.39$, $p = 0.019$, with the clinical entry participants having, on average, more childhood cross-gender behavior than the research entry participants. The percentage of clinical entry participants who were threshold for the diagnosis of GID in childhood did not differ significantly from the research entry participants (75.8 vs. 59.8%), $\chi^2_{(1)} = 1.83$. Of the 32 clinical entry participants, 8 had re-contacted the clinic because

[11] IQ data were not available for 11 of the 163 boys who were eligible for the study but were not contacted.

of gender dysphoria. The above-described comparisons were repeated to compare the research and clinical entry participants but with these 8 participants excluded. With the eight participants who contacted the clinic for gender dysphoria removed, there were no significant group differences on demographic variables, CBCL behavior problems, and measures of childhood sex-typed behavior (all $ps > 0.05$).

Gender Identity at Follow-Up

Appendix 2 in Supplementary Material shows the follow-up data for gender identity and sexual orientation for each participant. Of the 139 participants, 17 (12%) were classified as persisters and the remaining 122 (88%) were classified as desisters. The age at the time of follow-up did not differ significantly between the persisters (Mean, 20.12 years; SD = 5.54) and desisters (Mean, 20.64 years; SD = 5.19), $t_{(137)} < 1$. Of the 107 participants who, for research purposes only, were contacted for the follow-up study, 10 (9%) were classified as persisters; of the 32 participants who were recruited into the study after they were seen for some type of clinical concern, 7 (22%) were classified as persisters. The difference in persistence rate as a function of recruitment entry type was not significant, $\chi^2_{(1)} = 2.53$, $p = 0.112$. The difference in persistence rate between those patients seen for the face-to-face assessment vs. those who were not (14.5 vs. 3.4%) was also not significant, $\chi^2_{(1)} = 1.70$, $p = 0.192$. **Supplementary Table 1** summarizes information on some domains of gender role outcome for the 17 participants classified as having persistent gender dysphoria.

For the 42 participants where the GDIQ was used to determine gender identity status at follow-up, four were classified as persisters and 38 were classified as desisters. Of the 38 desisters, three endorsed one item and the remainder endorsed none of the items.[12] The four participants classified as persisters endorsed between three and five items.

For the 64 participants where the GIDYQ-AA was used to determine gender identity status at follow-up, 12 were classified as persisters and 52 were classified as desisters. All 52 desisters had a mean score >3.00 on the GIDYQ-AA. Of the 12 persisters, 10 had a mean score ≤3.00 and two had mean scores that were >3.00. In spite of having mean scores on the GIDYQ-AA that were above the recommended cutoff for caseness (95), these two participants were considered persisters because their clinical interview data indicated that they were experiencing significant gender dysphoria. Thus, clinical judgment was used to make the final classification for these two participants.

For the remaining 33 participants, clinical interview, parent-report or chart data were used to classify the percentage who were persisters ($n = 1$; 3%) or desisters ($n = 32$; 97%).

The persistence rate of gender dysphoria was examined as a function of participants' GID diagnostic status in childhood (threshold vs. subthreshold). Of the 88 participants who met the full diagnostic criteria for GID in childhood, 12 (13.6%) were classified as persisters and the remaining 76 (86.4%) were not. Of the 51 participants who were subthreshold for the GID diagnosis in childhood, 5 (9.8%) were classified as persisters and the remaining 46 (90.2%) were not. A chi-square analysis indicated that the rate of persistence did not differ significantly between the threshold and subthreshold groups, $\chi^2_{(1)} < 1$.

Over the years, prevalence rates for gender dysphoria in adults have varied considerably. The variation is likely a function of many factors, including definition, time period, and source of ascertainment. For example, in the Standards of Care of the World Professional Association for Transgender Health (112), probably relying on an estimate given in the DSM-IV-TR, the prevalence of gender dysphoria in adult males was estimated to be 1 in 30,000. In the meta-analysis by Arcelus et al. (113), the prevalence in adult males was estimated at 1 in 14,705. Lastly, Zhang et al.'s (114) review of recent population-based surveys estimated the prevalence of a self-reported transgender identity in adults to range between 0.33 and 0.53% (males and females combined). Regardless of which base rate figure one might choose to use as a point of comparison, the persistence rate of 12% (while low in an absolute sense) would be considerably higher than what one would detect in the general population.

Sexual Orientation at Follow-Up

Table 2 shows the Kinsey ratings for sexual orientation in fantasy. Data were not available for 10 participants, all of whom were desisters with regard to gender dysphoria. Based on the global rating for sexual orientation in fantasy, 43 (33.3%) participants were classified as gynephilic in fantasy and 82 (63.6%) were classified as biphilic/androphilic in fantasy. In the remaining four (3.1%) cases, the participants were classified as having no sexual fantasies and, therefore, a Kinsey rating could not be assigned.[13] In all four cases, the participants were desisters. Of the 17 participants classified as persisters, 1 (5.9%) was gynephilic in fantasy and 16 (94.1%) were biphilic/androphilic in fantasy. For participants assigned a Kinsey rating between 0 and 6 in fantasy, we correlated the interviewer's Kinsey rating with the participants' responses on the EROS in which their mean gynephilic score was subtracted from their mean androphilic score. This yielded an $r(101) = 0.86$, $p < 0.001$.

Table 2 also shows the Kinsey ratings for sexual orientation in behavior. Data were available for 108 participants. Based on the global rating for sexual orientation in behavior, 29 (26.9%) participants were classified as gynephilic and 51 (47.2%) were classified as biphilic/androphilic. The remaining 28 (25.9%) participants did not report any sexual behaviors in the 12 months preceding the follow-up assessment. For participants assigned a Kinsey rating between 0 and 6 in behavior, we correlated the

[12]By "endorsed," we mean that the participants answered other than "never" on Items 1–4 or response options d-e for Item 5 (see **Appendix 1** in Supplementary Material).

[13]For 104 participants, the Kinsey rating in fantasy was based on the information provided in the face-to-face interview. For 21 other participants, the Kinsey rating in fantasy was based on self-report (by telephone), information available in the participant's health record, or parent-report. Participants were assigned a Kinsey rating of 6 if the participant self-identified as "gay" or if the health record indicated that the patient was "homosexual" or gay, etc. Participants were assigned a Kinsey rating of 0 if the patient self-identified as "straight" or "heterosexual," etc. A chi-square test showed that the percentage of participants who were classified as Kinsey 0–1 vs. 2–6 did not differ significantly as a function sexual orientation ascertainment method, $\chi^2_{(1)} = 1.49$.

TABLE 2 | Kinsey ratings for sexual orientation in fantasy and behavior.

Variable	Kinsey rating (fantasy)[a]															
	0		1		2		3		4		5		6		No fantasy	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Crush	36	36.7	0	0	2	2.0	4	4.1	2	2.0	11	11.2	29	29.6	14	14.3
Visual	31	31.6	1	1.0	2	2.0	10	10.2	3	3.1	12	12.2	29	29.6	10	10.2
Dreams	13	13.3	1	1.0	1	1.0	4	4.1	3	3.1	3	3.1	27	27.6	46	46.9
Masturbation	21	21.9	2	2.1	3	3.1	6	6.3	2	2.1	7	7.3	33	34.4	22	22.9
Global fantasy rating	40	31.0	3	2.3	3	2.3	8	6.2	2	1.6	14	10.9	55	42.6	4	3.1
	Kinsey rating (behavior)[a]															
	0		1		2		3		4		5		6		No sexual behavior	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Holding hands	26	26.3	0	0	0	0	5	5.1	1	1.0	1	1.0	35	35.4	31	31.3
Kissing	21	21.2	0	0	0	0	6	6.1	2	2.0	2	2.0	34	24.3	34	34.3
Genital/breast contact	13	13.1	0	0	0	0	3	3.0	2	2.0	1	1.0	35	35.4	45	45.5
Intercourse	8	8.2	0	0	0	0	3	3.1	2	2.0	0	0	27	27.6	58	59.2
Global behavior rating	28	25.9	1	0.9	0	0	4	3.7	3	2.8	1	0.9	43	39.8	28	25.9

[a]*0 = Exclusively gynephilic to 6 = Exclusively androphilic.*

interviewer's Kinsey rating with the participants' responses on the SHQ in which their mean gynephilic score was subtracted from their mean androphilic score. This yielded an $r(75) = 0.79$, $p < 0.001$.

For those participants who could be assigned a Kinsey rating (i.e., excluding those participants who did not report any sexual fantasies or behavior or for whom data were not available), the correlation between Kinsey global fantasy and global behavior ratings was very strong, $r(78) = 0.92$, $p < 0.001$.

Group Classification as a Function of Gender Identity and Sexual Orientation in Fantasy at Follow-Up[14]

Combining gender identity (i.e., persister or desister) and sexual orientation in fantasy (i.e., gynephilic or biphilic/androphilic) at follow-up, the participants were classified into one of four outcome groups (for which we had all of the relevant data): (1) persistence of gender dysphoria with a biphilic/androphilic sexual orientation ($n = 16$); (2) desistance of gender dysphoria with a biphilic/androphilic sexual orientation ($n = 66$); (3) desistance of gender dysphoria with a gynephilic sexual orientation ($n = 42$); and (4) persistence of gender dysphoria with a gynephilic sexual orientation ($n = 1$). The participants who reported no sexual fantasies ($n = 4$) could not be included in this outcome classification. Given that only one participant was classified as gender dysphoric with a co-occurring gynephilic sexual orientation (Group 4), this category was excluded from subsequent analyses that compared these outcome groups.

[14]Given the strong correlation between Kinsey fantasy and behavior ratings and that there were fewer missing data on the Kinsey fantasy variable, participants were classified into one of the four outcome groups based on their fantasy ratings.

Demographic Characteristics in Childhood as a Function of Gender Identity and Sexual Orientation in Fantasy

Table 3 shows the demographic variables in childhood as a function of group. One-way ANOVAs and chi-square were conducted to evaluate whether the outcome groups differed on these variables. The groups differed significantly on four of the five childhood demographic variables. Duncan's multiple range test for unequal Ns showed that the biphilic/androphilic persisters were, on average, significantly older at the time of the childhood assessment than both the gynephilic desisters and the biphilic/androphilic desisters, who did not differ significantly from each other. The biphilic/androphilic desisters had, on average, a higher IQ than the biphilic/androphilic persisters but did not differ significantly from the gynephilic desisters. There was no significant difference in childhood IQ score between biphilic/androphilic persisters and gynephilic desisters. The biphilic/androphilic persisters were significantly more likely to come from a lower social class background compared to the gynephilic desisters and the biphilic/androphilic desisters, who did not differ significantly from each other (see also **Figure 1**). The biphilic/androphilic desisters were more likely to be living with both parents compared to the biphilic/androphilic persisters. There was no significant difference on marital status between the two desister groups.

The demographic variables from childhood on which the three groups differed–age at assessment, IQ, social class, and marital status–were significantly correlated (*r*s ranged from |0.32–0.58|) [see Table 12 in ("")]. To evaluate the predictive status of these variables on group outcome at follow-up, a multinomial logistic regression was performed. **Table 4** shows the results. For these analyses, the biphilic/androphilic desisters served as the reference

TABLE 3 | Demographic characteristics as a function of group.

Variable		Group			F or χ^2	p	η^2 or Cramer's V
		Persisters Biphilic/ Androphilic ($n = 16$)	Desisters Biphilic/ Androphilic ($n = 66$)	Desisters Gynephilic ($n = 42$)			
Childhood							
Age (in years)	*M*	6.85	6.96	7.49	3.57	0.031	0.06
	SD	1.67	2.69	2.62			
IQ[a]	*M*	101.63	110.20	103.18	3.77	0.026	0.06
	SD	14.81	14.56	15.16			
Social class[b]	*M*	23.76	44.97	39.44	15.30	<0.001	0.20
	SD	10.22	13.64	15.91			
Marital status[c]							
Two-parent	*N* (%)	7 (43.8)	49 (74.2)	24 (57.1)	6.74	0.034	0.23
Other	*N* (%)	9 (56.3)	17 (25.8)	18 (42.9)			
Ethnicity							
Caucasian	*N* (%)	14 (87.5)	58 (87.9)	32 (76.2)	2.77	0.250	0.14
Other	*N* (%)	2 (12.5)	8 (12.1)	10 (23.8)			
Follow-up							
Age at follow-up (in years)[d]	*M*	20.32	22.13	17.85	10.41	<0.001	0.15
	SD	5.67	4.97	3.95			
IQ at follow-up[a,e,f]	*M*	99.07	110.47	104.19	3.82	0.025	0.07
	SD	16.29	13.54	17.50			
Follow-up interval (in years)	*M*	11.47	15.17	10.36	9.63	<0.001	0.04
	SD	6.77	6.03	4.85			
Social desirability[g]	*M*	0.44	0.43	0.52	3.07	0.051	0.07
	SD	0.17	0.16	0.19			

[a] *Full-Scale IQ was obtained with age-appropriate Wechsler intelligence scales.*
[b] *Hollingshead's (1975) Four Factor Index of Social Status (absolute range, 8–66).*
[c] *Other included the following family constellations: single parent, separated, divorced, living with relatives, or in the care of a child protection agency.*
[d] *Interval denotes the time between childhood assessment and follow-up assessment.*
[e] *Full Scale IQ was estimated using four subtests: Vocabulary, Comprehension, Block Design, and Object Assembly.*
[f] *An IQ score was available only for participants who completed the face-to-face assessment.*
[g] *Absolute range, 0.00–1.00. Higher score indicates a greater propensity to give socially desirable responses. Age at follow-up, IQ at follow-up, social class, and parent's marital status were co-varied.*

group. Each coefficient, *B*, represents the change in the log odds for Group for a 1-unit increase in the corresponding predictor, controlling for all other predictors in the model. The next column presents the standard error (SE) for each *B*. The Wald statistic was the quantity used to determine the significance level of each predictor variable. The quantity, e^B, is the multiplicative change in the odds of being classified as a biphilic/androphilic persister (Model 1) or a gynephilic desister (Model 2) for a 1-unit increase in the corresponding predictor, and thus $100 \times (e^B - 1)$ represents the percentage change in the odds ratio for a 1-unit increase in that predictor (115).

It can be seen from **Table 4** that only social class had a significant contribution to the prediction of group outcome at follow-up (see also **Figure 1**). The biphilic/androphilic persisters had a 13% increase in odds of coming from a lower social class background compared to the biphilic/androphilic desisters. However, social class did not predict outcome when the two desister groups were compared.

Table 3 also shows the variables of age, IQ, and social desirability scores at follow-up as a function of group. One-way ANOVAs revealed that both age and IQ differed significantly among the three groups ($ps < 0.01$), but social desirability scores did not. Duncan's multiple range test for unequal Ns showed that the gynephilic desisters were, on average, younger than both the biphilic/androphilic persisters and the biphilic/androphilic desisters (both $ps < 0.05$), who did not differ significantly from each other. Regarding IQ at follow-up, the results were similar to those for IQ in childhood. The biphilic/androphilic desisters had, on average, a higher IQ than the biphilic/androphilic persisters ($p < 0.05$) but did not differ significantly from the gynephilic desisters. There was no significant difference in IQ between the biphilic/androphilic persisters and the gynephilic desisters.

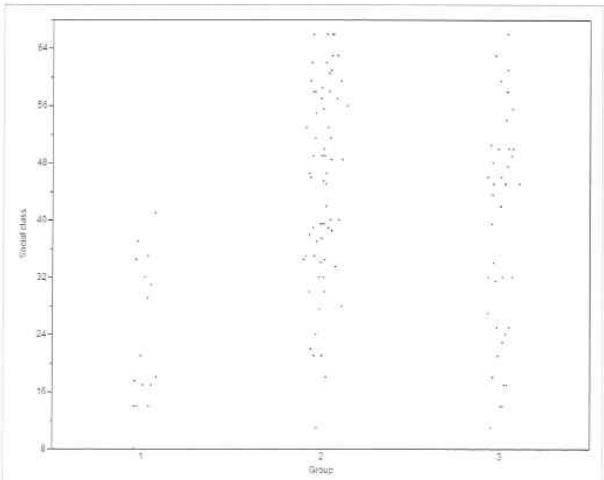


FIGURE 1 | Distribution of social class for the outcome groups at follow-up. 1 = Biphilic/androphilic persisters (n = 16; M = 23.76, SD = 10.22), 2 = Biphilic/androphilic desisters (n = 66; M = 44.97, SD = 13.64), 3 = Gynephilic desisters (n = 42; M = 39.44, SD = 15.91).

Childhood Sex-Typed Behavior as a Function of Gender Identity and Sexual Orientation at Follow-Up

Supplementary Table 2 shows the means or percentage scores (for dichotomous measures) of the nine sex-typed measures obtained at the assessment in childhood as a function of the three outcome groups. ANCOVAs (with age at assessment, IQ, social class, and marital status covaried) or chi-square were used to examine whether the groups differed on any of these variables.[15] There was a significant difference between the groups on four child-report measures (first drawn person on the Draw-a-Person, free play, Gender Identity Interview, and cross-sex peer preference on the Playmate and Play Style Preferences Structured Interview, and one parent-report measure (Gender Identity Questionnaire for Children). A statistical summary of these individual measures can be found in the **Supplementary Text** and the data are shown in **Supplementary Table 2**.

The childhood sex-typed behavior measures on which the groups differed were all significantly correlated (rs ranged from |0.30–0.76|) [reported in (77), Table 15].[16] From these six measures (first drawn person on the Draw-a-Person, free play, Gender Identity Interview, cross-sex peer preference on the Playmate and Play Style Preferences Structured Interview, cross-sex toy preference on the Playmate and Play Style Preferences Structured Interview, and the Gender Identity Questionnaire for Children), a composite score of childhood sex-typed behavior was derived for each participant by taking the average of the six variables (each expressed as z-scores).[17] A higher composite z-score indicates more cross-gender behavior at the assessment in childhood.

To evaluate the influence of childhood sex-typed behavior and demographic variables on group outcome at follow-up, a multinomial logistic regression was performed using the composite score and the demographic variables on which the groups differed–age at assessment, IQ, and social class–as predictor variables. It can be seen from **Table 5** that both social class and the composite score of childhood sex-typed behavior were significant predictors of group outcome at follow-up in the first model, which compared the biphilic/androphilic persisters to the biphilic/androphilic desisters.

The biphilic/androphilic persisters had a 274% increase in odds of having a higher composite score (i.e., more childhood cross-gender behavior) and an 11% reduction in the odds of coming from a higher social class compared to the biphilic/androphilic desisters. Age at childhood assessment and IQ did not have a significant effect on group outcome (both ps > 0.05). In the second model, which compared the gynephilic desisters to the biphilic/androphilic desisters, the only significant predictor of group outcome was the composite measures of sex-typed behavior. The biphilic/androphilic desisters had a 48% increase in odds of having a higher composite score compared to the gynephilic desisters.

DISCUSSION

Methodological Issues

We were not able to recruit into the study all eligible patients; however, our analyses which compared the participants vs. the non-participants did not show any substantive or pervasive differences with regard to the baseline assessment characteristics, suggesting that the internal validity of the sample was not grossly compromised (111). The majority of follow-up participants were recruited for research purposes; however, a minority entered the study after having been seen in adolescence for some clinical issue. There was some evidence that the patients who were enrolled in the study after recontacting the clinic were, on average, more extreme in their gender-variant behavior in childhood; however, the percentage who were threshold for the GID diagnosis in childhood did not differ significantly between the two subgroups. Although the percentage of persisters was higher in the subgroup that had recontacted the clinic than the subgroup recruited for research purposes only (22% vs. 9%), the difference was also not statistically significant. If anything, the direction of the difference would suggest that the overall rate of persistence may have been slightly overestimated had we relied entirely on a "research-only" follow-up sample.

Another methodological issue is that we relied on different metrics to assess gender identity and gender dysphoria at follow-up. For example, we replaced the GDIQ with the GIDYQ-AA as we viewed the latter as a better measure; in some instances,

[15] The ANCOVA model was adjusted to accommodate a categorical covariate.

[16] Although the groups did not differ significantly on cross-sex toy preference on the PPPSI, this measure is included here because there was a trend in the direction of a significant group difference.

[17] For some participants, data were not available on all six measures. In these cases, the composite score was the average of the number of variables for which there were data.

TABLE 4 | Multinomial logistic regression of group outcome at follow-up.

Predictor	Biphilic/Androphilic persisters					Gynephilic desisters				
	B	SE	Wald	*p*	e^B	*B*	SE	Wald	*p*	e^B
Age at assessment	0.11	0.14	0.62	0.433	1.12	−0.02	0.09	0.03	0.856	0.98
IQ	0.02	0.03	0.85	0.358	1.02	−0.02	0.02	1.91	0.167	0.98
Social class	−0.14	0.04	13.66	<0.001	0.87	−0.01	0.02	0.13	0.716	0.99
Marital status	0.76	0.80	0.88	0.349	0.47	−0.43	0.52	0.70	0.402	1.54

Reference group is the Biphilic/Androphilic Desisters. This group was chosen as the reference because it had the largest group size.

TABLE 5 | Multinomial logistic regression predicting group outcome at follow-up.

Predictor	Biphilic/Androphilic persisters					Gynephilic desisters				
	B	SE	Wald	*p*	e^B	*B*	SE	Wald	*p*	e^B
Age at assessment	0.26	0.16	2.90	0.09	1.30	−0.14	0.11	1.55	0.21	0.87
IQ	0.02	0.03	0.58	0.45	1.02	−0.03	0.01	2.77	0.10	0.97
Social class	−0.12	0.03	12.28	<0.001	0.89	−0.01	0.01	0.51	0.47	0.99
Composite *z*-score	1.32	0.55	5.82	0.02	3.74	−0.66	0.31	4.38	0.04	0.52

Reference group is the Biphilic/Androphilic Desisters. This group was chosen as the reference because it had the largest group size. A preliminary analysis with marital status included as a predictor variable showed that it did not have a significant effect and was, therefore, excluded in the final regression model. As suggested by Reviewer 3, per Benjamin et al. (116), for the "discovery of new effects," p-values between 0.05 and 0.005 should be viewed as "suggestive" (i.e., informative, but cautiously interpreted), and p-values < 0.005 as "significant" (i.e., stronger evidence for the implausibility of a difference merely by chance).

we relied solely on interview data or information available in the patient's medical chart. However, we did not detect any substantive difference in the percentage of persisters across these different sources of information and thus do not believe that such method variance challenges the validity of the findings.

Although a minority of participants were seen on more than one occasion for follow-up, the majority were not. Thus, our results and interpretation of the follow-up data are largely limited to one "moment in time," at a mean age of 20.58 years. It would, of course, be of value to have additional follow-up of the patients as they move further into adulthood in order to assess the stability (or lack thereof) of the data with regard to both gender identity and sexual orientation. In our own clinical experience, for example, we have observed that some of the patients seen during adolescence "fluctuated" between self-identifying as transgender and self-identifying as gay. Others have noted that a small number of apparent or presumed desisters during adolescence subsequently identified as transgender when seen at a later point in time (117).

Summary of Key Findings

The present study provided follow-up data with regard to gender identity and sexual orientation in boys referred clinically for gender dysphoria. There were three key findings: (1) the persistence of gender dysphoria was relatively low (at 12%), but obviously higher than what one would expect from base rates in the general population; (2) the percentage who had a biphilic/androphilic sexual orientation was very high (in fantasy: 65.6% after excluding those who did not report any sexual fantasies; in behavior: 63.7% after excluding those who did not have any interpersonal sexual experiences), markedly higher than what one would expect from base rates in the general population; (3) we identified some predictors (from childhood) of long-term outcome when contrasting the persisters with a biphilic/androphilic sexual orientation with the desisters with a biphilic/androphilic sexual orientation and when contrasting the desisters with a biphilic/androphilic sexual orientation and the desisters with a gynephilic sexual orientation.

The 12% persistence rate was somewhat lower than the overall persistence rate of 17.4% from the prior follow-up studies of boys combined. When compared to the three most methodologically sound follow-up studies, the persistence rate was higher than the 2.2% rate found by Green (47), but lower than the 20.3% rate found by Wallien and Cohen-Kettenis (52) and the 29.1% rate found by Steensma et al. (51). There is one methodological caveat regarding the Steensma et al. study that is worth noting. In their study, the mean interval between assessment and follow-up was relatively short (7.21 years). The patients were eligible for follow-up if they were at least 15 years of age. Given the relatively short interval between the assessment in childhood and the follow-up assessment in adolescence, this meant that patients who had been assessed at younger ages in childhood would not have been old enough to participate in the follow-up assessment. Given that Steensma et al. found that (older) age at the time of the assessment in childhood was a significant predictor of persistence, it is conceivable that their persistence rate was an overestimate. Nonetheless, in the broadest sense, our data were quite consistent with the general finding from the prior follow-up studies that desistance from gender dysphoria is by far the more common outcome.

In our study, we did not find that persistence was more common among boys who were threshold for the diagnosis of GID when compared to the boys who were subthreshold (13.6% vs. 9.8%) although the pattern was in the same direction

as that found by Wallien and Cohen-Kettenis (52) and Steensma et al. (51). We would, therefore, argue that the threshold-subthreshold distinction should not be abandoned in future follow-up studies although such studies might profit from using a symptom count of DSM indicators in addition to the dichotomous coding of the diagnosis as threshold vs. subthreshold. Consistent with both Wallien and Cohen-Kettenis and Steensma et al., our composite measure of sex-typed behavior in childhood was a significant predictor of outcome in that the patients classified as persisters with a biphilic/androphilic sexual orientation had more severe gender-variant behavior than the patients classified as desisters with a biphilic/androphilic sexual orientation; in addition, desisters with a biphilic/androphilic sexual orientation had more gender-variant behavior than the desisters with a gynephilic sexual orientation. Thus, dimensional measurement of gender identity and gender role behaviors from childhood provides added nuance in characterizing longer term trajectories with regard to both gender identity and sexual orientation.

With regard to sexual orientation at follow-up, the percentage of patients with a biphilic/androphilic sexual orientation in either fantasy or behavior was reasonably similar to those reported on in the prior follow-up studies which included standardized assessment measures (47, 51, 52). This finding also converges with three representative, general population prospective studies (118–120) and many retrospective studies (43) which document a significant association between patterns of gender-typed behavior in childhood and later sexual orientation.

The multinomial logistic regression analysis (**Table 4**) also showed a trend for the persisters with a biphilic/androphilic sexual orientation to be older at the time of the assessment in childhood compared to the desisters with a biphilic/androphilic sexual orientation; however, when the composite measure of sex-typed behavior in childhood was added to the equation (**Table 5**), age at assessment in childhood no longer showed such a trend [cf. Steensma et al. (51)]. In our smaller study of girls with GID (46), the persisters were, on average, 2.5 years older than the desisters at the time of the assessment in childhood (11.08 vs. 8.59 years) although the difference was not significant. It is our view that age at the time of a childhood assessment in relation to long-term outcome should continue to be examined in future follow-up studies.

Social class was a significant predictor of outcome: the persisters with a biphilic/androphilic sexual orientation were from a lower social class background compared to the desisters with a biphilic/androphilic sexual orientation (even after controlling for the other demographic variables). Why might this be the case? Because we had not made formal a priori predictions of outcome regarding any of our demographic variables, it is, of course, important to see whether or not it will be replicated in new follow-up studies. At present, our interpretation of the social class effect reflects on its relationship to other literatures.

One possibility pertains to the notion that acceptance of a gay or homosexual sexual identity is less in "working class" subculture (121). If this is, in fact, the case, it has been argued that transitioning from male to female—the so-called "homophobic" hypothesis with regard to gender dysphoria in adults (122)–would allow an androphilic sexual orientation to be more acceptable. Future studies would need to systematically examine whether boys with persistent GID first attempt to live as gay men before transitioning to the female gender role and whether or not this temporal sequence, when it occurs, is related to social class background.

In the present study, it could be hypothesized that the parents of persisters held less favorable views of androphilia (homosexuality) compared to the desisters and thus predisposed to persistence in order to "normalize" one's sexual orientation. However, this is simply a conjecture as parental attitudes toward homosexuality were not measured in the study sample. Indeed, none of the follow-up studies to date on boys with gender dysphoria have specifically examined attitudes toward homosexuality as a predictor of outcome.

Social class could also be a proxy for other explanatory factors. For example, in the present study, a lower social class background was significantly correlated with age at assessment in childhood ($r = 0.44$) and families where there had been a separation/divorce, etc. ($r = 0.58$). If one wanted to make the case that a later age at assessment might be associated with persistence (for a variety of reasons), perhaps social class is associated with a "delay" in seeking out an assessment and possible treatment (e.g., family stress, various other mental health challenges in the child and/or the family, etc.). In one study comparing the demographic characteristics of children vs. adolescents clinic-referred for gender dysphoria, it was found that the adolescents were more likely than the children to come from a lower social class background and from families in which there had been a separation/divorce, etc. (123).

Clinical Implications

What clinical implications might be drawn from our data on the persistence and desistence rates of gender dysphoria in children? First, it should be recognized that the boys in the current study were seen during a period of time when treatment recommendations, if such were made, often aimed to reduce the gender dysphoria between the child's felt gender identity and biological sex. If one peruses the treatment literature, such recommendations were carried out using many therapeutic modalities: psychotherapy or psychoanalysis, behavior therapy, group therapy, parent-counseling, and interventions in the naturalistic environment, such as encouragement of same-sex peer relations [see, e.g., (124–126); for reviews, see (127, 128)].[18]

[18]This "broad stroke" summary of therapeutic goals is not meant to minimize the complexity of ethical issues regarding how treatment has been conceptualized over the years [see, e.g., (129–133)]. In the early years, treatment recommendations included other goals: for example, Bakwin (44) wrote that "Suggestions for management…[were]…designed to encourage gender appropriate behavior *and* to prevent homosexuality" [p. 620, emphasis added; see also (134)]. Rekers (135) was subsequently quite transparent regarding the influence of his own religious beliefs in formulating treatment goals, sometimes congruent with parents' religious beliefs (see p. 131). Prayer appears to have guided Rekers' selection of behavior therapy as a treatment modality for the treatment of his patients with childhood GID (p. 131). Money and Russo (50) wondered what the course of psychosexual differentiation might be if "a group of boys with discordance of gender identity/role [were] transferred from the home of origin to, say, a children's recovery center or foster home…as happens in the case of child-abuse dwarfism…" (p. 40). In our own clinic, although some parents might have desired or requested that treatment be designed in order to prevent homosexuality, this was a goal that we never endorsed [see (136), pp. 391–393]. Over the years, many secular-minded

In our own sample, the kinds of treatments that the boys received, if any, were quite variable but it is beyond the scope of this article to describe them in general [however, for examples, see (136, 140, 141)]. It can, however, be said with certainty that the vast majority of boys were seen during a particular period of time when the therapeutic approach of recommending or supporting a gender social transition prior to puberty was not made. Indeed, in the current study, there was only one patient who had socially transitioned prior to puberty (at the suggestion and support of the professionals involved in this individual's care) and this particular patient was one of the persisters with a biphilic/androphilic sexual orientation. Second, it should also be recognized that, for the boys seen in the current study, none who were in late childhood and had (likely) entered puberty (Tanner Stage 2) had received puberty-blocking hormone treatment (GnRH analogs) to suppress somatic masculinization (142, 143) until sometime during adolescence.

In contrast, in recent years, it has become more common for some clinicians to recommend a gender social transition prior to puberty [e.g., (69, 144–147); for discussion, see (148–150)]. It has also become more common for parents to have already implemented a gender social transition on their own, without any formal input from a health professional (151). As argued by Zucker (64, 152), this is a very different type of psychosocial treatment designed to reduce gender dysphoria when compared to the other kinds of treatments noted above that have been recommended over the years.

The study by Steensma et al. (51), which found the highest rate of persistence, included some patients who had made a partial or complete gender social transition prior to puberty and this variable proved to be a unique predictor of persistence (see the Introduction). Rae et al. (153) recruited from a variety of community groups a sample of 85 markedly gender non-conforming children (Mean age, 7.5 years), none of whom had socially transitioned at a baseline assessment. At the time of follow-up, at a mean of 2.1 years later, 36 (42.3%) had socially transitioned and 49 (57.6%) had not. Using a composite of various metrics of gender identity and gender role behaviors, Rae et al. found that those who subsequently socially transitioned had more extreme gender-variant behavior at baseline than those who had not. Thus, this short-term follow-up study was consistent with the longer-term findings reported on by Wallien and Cohen-Kettenis (52), Steensma et al. (51), and the present study.

To date, however, there are no long-term follow-up studies of clinic-referred samples of children who had all socially transitioned prior to puberty. Future follow-up studies should be able to capture a much larger subgroup of such children and compared to those who have not with regard to long-term outcome with regard to persistence and desistance [e.g., (154)]. The persistence-desistance rates found in this study and the ones preceding it can be used as a comparative benchmark for samples in which a social transition took place prior to puberty.

clinicians–although clearly opposed to any type of preventive efforts with regard to sexual orientation–argued in favor of reducing gender dysphoria vis-à-vis natal sex, if that was feasible. Meyer-Bahlburg (125), for example, wrote: "…we cannot rule out the possibility that early successful treatment of childhood GID will diminish the role of a continuation of GID into adulthood. If so, successful treatment would also reduce the need for the long and difficult process of sex reassignment which includes hormonal and surgical procedures with substantial medical risks and complications" (p. 362). Along similar lines, Cohen-Kettenis and Pfäfflin (33) remarked: "Relatively little dispute exists regarding the prevention of transsexualism, though evidence about the effectiveness of treatment in preventing adult transsexualism is also virtually nonexistent" (p. 120). In more recent years, what the best-practice should be for the treatment of gender dysphoria in children has been widely discussed and debated, which highlight the various limitations of treatment effectiveness studies (137–139).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The research protocol was reviewed and approved by Clarke Institute of Psychiatry (subsequently the Centre for Addiction and Mental Health) and the University of Toronto. All participants who completed the face-to-face assessment gave written informed consent.

AUTHOR CONTRIBUTIONS

DS contributed to the conceptualization, data collection, data analysis, interpretation, and writing of the paper. SB contributed to the conceptualization and interpretation of the study. KZ contributed to the conceptualization, data collection, data analysis, interpretation, and writing of the paper. All authors contributed to the article and approved the submitted version.

FUNDING

DS was supported by an Ontario Graduate Scholarship (2008–2009, 2009–2010, 2010–2011) and the Social Sciences Humanities Research Council (2010–2011). Funding for this study was provided, in part, by the Laidlaw Foundation and internal research funds from the Clarke Institute of Psychiatry.

ACKNOWLEDGMENTS

This article is based on DS's doctoral dissertation at the University of Toronto. Preliminary versions of this article were presented at the 1989 meeting of the Society for Research in Child and Adolescent Psychopathology, Miami Beach, Florida and the 2010 Gender Development Research Conference, San Francisco, California.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.632784/full#supplementary-material

REFERENCES

1. Lewis M. Self-knowledge: a Social cognitive perspective on gender identity and sex-role development. In: Lamb ME, Sherrod LR, editors. *Infant Social Cognition: Empirical and Theoretical Considerations*. Hillsdale, NJ: Lawrence Erlbaum Associates (1981). p. 395–414.
2. Maccoby EE. Gender as a social category. *Dev Psychol.* (1988) 24:755–65. doi: 10.1037/0012-1649.24.6.755
3. Ruble DN, Martin CL, Berenbaum SA. Gender development. In: Damon W, Lerner RM, Eisenberg M, editors. *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality Development*, 6th ed. New York, NY: Wiley (2006). p. 858–932.
4. Stoller RJ. The sense of maleness. *Psychoanal Q.* (1965) 34:207–18. doi: 10.1080/21674086.1965.11926345
5. Stoller RJ. The sense of femaleness. *Psychoanal Q.* (1968) 37:42–55. doi: 10.1080/21674086.1968.11926450
6. Zucker KJ, VanderLaan DP. The self in gender dysphoria: a developmental perspective. In: Kyrios M, Moulding R, Doron G, Bhar SS, Nedeljkovic M, Mikulincer M, editors. *The Self in Understanding and Treating Psychological Disorders*. Cambridge: Cambridge University Press (2016). p. 222–32. doi: 10.1017/CBO9781139941297.023
7. Turner KL, Brown CS. The centrality of gender and ethnic identities across individuals and contexts. *Soc Dev.* (2007) 16:701–19. doi: 10.1111/j.1467-9507.2007.00403.x
8. Matei A. Raising a Theybie: The Parent Who Wants Their Child to Grow Up Gender-Free. *The Guardian* (2020). Retrieved from: https://www.theguardian.com/lifeandstyle/2020/jul/08/parent-raising-gender-free-child
9. Myers K. *Raising Them: Our Adventure in Gender Creative Parenting*. New York, NY: TOPPLE Books/Little A (2020).
10. Witterick K. Dancing in the eye of the storm: the gift of gender diversity to our family. In: Green FG, Friedman M, editors. *Chasing Rainbows: Exploring Gender Fluid Parenting Practices*. Bradford, ON: Demeter Press (2013). p. 21–42.
11. Kleeman JA. The establishment of core gender identity in normal girls. I (a) Introduction; (b) Development of the ego capacity to differentiate. *Archiv Sexual Behav.* (1971) 1:103–16. doi: 10.1007/BF01541055
12. Martin CL, Ruble DN, Szkrybalo J. Cognitive theories of early gender development. *Psychol Bull.* (2002) 128:903–33. doi: 10.1037/0033-2909.128.6.903
13. Paluszny M, Beit-Hallahmi B, Catford JC, Cooley RE, Dull CY, Guiora AZ. Gender identity and its measurement in children. *Compreh Psychiatry.* (1973) 14:281–90. doi: 10.1016/S0010-440X(73)80022-7
14. Slaby RG, Frey KS. Development of gender constancy and selective attention to same-sex models. *Child Dev.* (1975) 46:849–56. doi: 10.2307/1128389
15. Kohlberg L. A cognitive-developmental analysis of children's sex-role concepts and attitudes. In: Maccoby EE, editor. *The Development of Sex Differences*. Stanford, CA: Stanford University Press (1966). p. 82–173.
16. Zucker KJ, Bradley SJ, Kuskis M, Pecore K, Birkenfeld-Adams A, Doering RW, et al. Gender constancy judgments in children with gender identity disorder: evidence for a developmental lag. *Archiv Sexual Behav.* (1999) 28:475–502. doi: 10.1023/A:1018713115866
17. Fagot BI, Leinbach MD, Hagan R. Gender labelling and adoption of sex-typed behaviors. *Dev Psychol.* (1986) 22:440–3. doi: 10.1037/0012-1649.22.4.440
18. Zucker KJ. Measurement of psychosexual differentiation. *Archiv Sexual Behav.* (2005) 34:375–88. doi: 10.1007/s10508-005-4336-7
19. Hines M. *Brain Gender*. Oxford: Oxford University Press (2004).
20. Owen Blakemore JE, Berenbaum SA, Liben LS. *Gender Development*. New York, NY: Taylor & Francis Group (2009).
21. Weisgram ES, Dinella LM, editors. *Gender Typing of Children's Toys: How Early Play Experiences Impact Development*. Washington, DC: American Psychological Association (2018).
22. Lippa RA. Gender-related traits in gay men, lesbian women, and heterosexual men and women: the virtual identity of homosexual-heterosexual diagnosticity and gender diagnosticity. *J Pers.* (2000) 68:899–926. doi: 10.1111/1467-6494.00120
23. Bakwin H. Transvestism in children. *J Pediatr.* (1960) 56:294–8. doi: 10.1016/S0022-3476(60)80128-X
24. Friend MR, Schiddel L, Klein B, Dunaeff D. Observations on the development of transvestitism in boys. *Am J Orthopsychiatry.* (1954) 24:563–75. doi: 10.1111/j.1939-0025.1954.tb06128.x
25. Green R, Money J. Incongruous gender role: nongenital manifestations in prepubertal boys. *J Nerv Ment Dis.* (1960) 131:160–8. doi: 10.1097/00005053-196008000-00009
26. Stoller RJ. Male childhood transsexualism. *J Am Acad Child Psychiatry.* (1968) 7:193–209. doi: 10.1016/S0002-7138(09)62167-1
27. Zuger B. Effeminate behavior present in boys from early childhood I. The clinical syndrome and follow-up studies. *J Pediatr.* (1966) 69:1098–107. doi: 10.1016/S0022-3476(66)80301-3
28. Stoller RJ. *Sex and Gender (Vol. I). The Development of Masculinity and Femininity*. New York, NY: Jason Aronson (1968).
29. Green R. *Sexual Identity Conflict in Children and Adults*. New York, NY: Basic Books (1974).
30. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed. Washington, DC: American Psychiatric Association (1980).
31. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Arlington, VA: American Psychiatric Press (2013).
32. Coates S. Ontogenesis of boyhood gender identity disorder. *J Am Acad Psychoanal.* (1990) 18:414–38. doi: 10.1521/jaap.1.1990.18.3.414
33. Cohen-Kettenis PT, Pfäfflin F. *Transgenderism and Intersexuality in Childhood and Adolescence: Making Choices*. Thousand Oaks, CA: Sage (2003).
34. de Vries ALC, Kreukels BPC, Steensma TD, McGuire JK. Gender identity development: a biopsychosocial perspectives. In: Kreukels BPC, Steensma TD, de Vries ALC, editors. *Gender Dysphoria and Disorders of Sex Development: Progress in Care and Knowledge*. New York, NY: Springer (2014). p. 53–80. doi: 10.1007/978-1-4614-7441-8_3
35. Spivey LA, Edwards-Leeper L. Future directions in affirmative psychological interventions with transgender children and adolescents. *J Clin Child Adolesc Psychol.* (2019) 48:343–56. doi: 10.1080/15374416.2018.1534207
36. Turban JL, Ehrensaft D. Gender identity in youth: treatment paradigms and controversies. *J Child Psychol Psychiatry.* (2018) 59:1228–43. doi: 10.1111/jcpp.12833
37. Zucker KJ. Gender identity disorder in children and adolescents. *Annu Rev Clin Psychol.* (2005) 1:467–92. doi: 10.1146/annurev.clinpsy.1.102803.144050
38. Zucker KJ. Gender dysphoria. In: Lewis M, Rudolph KD, editors. *Handbook of Developmental Psychopathology*, 3rd ed. New York, NY: Springer (2014). p. 683–702. doi: 10.1007/978-1-4614-9608-3_35
39. Zucker KJ, Bradley SJ. *Gender Identity Disorder and Psychosexual Problems in Children and Adolescents*. New York, NY: Guilford Press (1995).
40. Green R. The behaviorally feminine male child: pretranssexual? Pretransvestic? Prehomosexual? Preheterosexual? In: Friedman RC, Richart RM, Vande Wiele RL, editors. *Sex Differences in Behavior*. New York, NY: John Wiley & Sons (1974). p. 301–14.
41. Bell AP, Weinberg MS, Hammersmith SK. *Sexual Preference: Its Development in Men and Women*. Bloomington: Indiana University Press (1981).
42. Whitam F. Childhood indicators of male homosexuality. *Archiv Sexual Behav.* (1977) 6:89–96. doi: 10.1007/BF01541701
43. Bailey JM, Zucker KJ. Childhood sex-typed behavior and sexual orientation: a conceptual analysis and quantitative review. *Dev Psychol.* (1995) 31:43–55. doi: 10.1037/0012-1649.31.1.43
44. Bakwin H. Deviant gender-role behavior in children: relation to homosexuality. *Pediatrics.* (1968) 41:620–9.
45. Davenport CW. A follow-up study of 10 feminine boys. *Archiv Sexual Behav.* (1986) 15:511–7. doi: 10.1007/BF01542316
46. Drummond KD, Bradley SJ, Peterson-Badali M, Zucker KJ. A follow-up study of girls with gender identity disorder. *Dev Psychol.* (2008) 44:34–45. doi: 10.1037/0012-1649.44.1.34
47. Green R. *The "Sissy Boy Syndrome" and the Development of Homosexuality*. New Haven, CT: Yale University Press (1987).
48. Kosky RJ. Gender-disordered children: does inpatient treatment help? *Med J Aust.* (1987) 146:565–9. doi: 10.5694/j.1326-5377.1987.tb120415.x
49. Lebovitz PS. Feminine behavior in boys: aspects of its outcome. *Am J Psychiatry.* (1972) 128:1283–9. doi: 10.1176/ajp.128.10.1283

50. Money J, Russo AJ. Homosexual outcome of discordant gender identity/role in childhood: longitudinal follow-up. *J Pediatr Psychol.* (1979) 4:29–41. doi: 10.1093/jpepsy/4.1.29
51. Steensma TD, McGuire JK, Kreukels BPC, Beekman AJ, Cohen-Kettenis PT. Factors associated with desistence and persistence of childhood gender dysphoria: A quantitative follow-up study. *J Am Acad Child Adolesc Psychiatry.* (2013) 52:582–90. doi: 10.1016/j.jaac.2013.03.016
52. Wallien MSC, Cohen-Kettenis PT. Psychosexual outcome of gender dysphoric children. *J Am Acad Child Adolesc Psychiatry.* (2008) 47:1413–23. doi: 10.1097/CHI.0b013e31818956b9
53. Zuger B. Effeminate behavior present in boys from childhood: ten additional years of follow-up. *Compreh Psychiatry.* (1978) 19:363–9. doi: 10.1016/0010-440X(78)90019-6
54. August GJ, Realmuto GM, Joyce T, Hektner JM. Persistence and desistance of oppositional defiant disorder in a community sample of children with ADHD. *J Am Acad Child Adolesc Psychiatry.* (1999) 38:1262–70. doi: 10.1097/00004583-199910000-00015
55. Zucker KJ. Persistence and desistance of gender identity disorder in children [Discussant]. In: *Paper Presented at the Meeting of the Harry Benjamin International Gender Dysphoria Association*, Gent, Belgium (2003).
56. Cantor JM. Transgender and gender diverse children and adolescents: fact-checking of AAP policy. *J Sex Marit Ther.* (2020) 46:307–13. doi: 10.1080/0092623X.2019.1698481
57. Ristori J, Steensma TD. Gender dysphoria in childhood. *Int Rev Psychiatry.* (2016) 28:13–20. doi: 10.3109/09540261.2015.1115754
58. Zuger B. Early effeminate behavior in boys: outcome and significance for homosexuality. *J Nerv Ment Dis.* (1984) 172:90–7. doi: 10.1097/00005053-198402000-00005
59. Nakamura H. *Follow-up study of children and adolescents with gender identity development issues who attended the specialist Gender Identity Development Unit (GIDU) and who are now 18 or older* (Unpublished dissertation), University of Essex, Essex, United Kingdom (2007).
60. Rafferty J. Committee on Psychosocial Aspects of Child and Family Health, Committee on Adolescence, and Section on Lesbian, Gay, Bisexual, and Transgender Health and Wellness. Ensuring comprehensive care and support for transgender and gender-diverse children and adolescents. *Pediatrics.* (2018) 142:e20182162. doi: 10.1542/peds.2018-2162
61. Tannehill B. The End of the Desistance Myth. *Huffpost* (2016). Retrieved from: https://www.huffingtonpost.com/brynn-tannehill/the-end-of-the-desistance_b_8903690.html
62. Temple Newhook J, Pyne J, Winters K, Feder S, Holmes C, Tosh J, et al. A critical commentary on follow-up studies and "desistance" theories about transgender and gender-nonconforming children. *Int J Transgen.* (2018) 19:212–24. doi: 10.1080/15532739.2018.1456390
63. Winters K. The "80% desistance dictum: is it science? In: Lev AI, Gottlieb AR, editors. *Families in Transition: Parenting Gender Diverse Children, Adolescents, and Young Adults*. New York, NY: Harrington Park Press (2019). p. 88–101.
64. Zucker KJ. The myth of persistence: Response to "A Critical Commentary on Follow-Up Studies and Desistance Theories about Transgender and Gender Non-Conforming Children" by Temple Newhook et al. (2018). *Int J Transgen.* (2018) 19:231–45. doi: 10.1080/15532739.2018.1468293
65. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed. rev. Washington, DC: American Psychiatric Association (1987).
66. Zucker KJ. The DSM diagnostic criteria for gender identity disorder in children. *Archiv Sexual Behav.* (2010) 39:477–98. doi: 10.1007/s10508-009-9540-4
67. Zucker KJ, Cohen-Kettenis PT, Drescher J, Meyer-Bahlburg HFL, Pfäfflin F, Womack WM, Memo outlining evidence for change for gender identity disorder in the DSM-5. *Archiv Sexual Behav.* (2013) 42:901–14. doi: 10.1007/s10508-013-0139-4
68. Fast AA, Olson KR. Gender development in transgender preschool children. *Child Dev.* (2018) 89:620–37. doi: 10.1111/cdev.12758
69. Meadow T. *Trans Kids: Being Gendered in the Twenty-First Century*. Berkeley, CA: University of California Press (2018).
70. Olson KR. Prepubescent transgender children: what we do and do not know. *J Am Acad Child Adolesc Psychiatry.* (2016) 55:155–6. doi: 10.1016/j.jaac.2015.11.015
71. Soares C, editor. *The Oxford Dictionary of Current English*, 3rd ed. Oxford: Oxford University Press (2001).
72. Blanchard R, Zucker KJ, Cavacas A, Allin S, Bradley SJ, Schachter DC. Fraternal birth order and birth weight in probably prehomosexual feminine boys. *Horm Behav.* (2002) 41:321–7. doi: 10.1006/hbeh.2002.1765
73. Bouman WP, Suess Schwend A, Motmans J, Smily A, Safer JD, Deutsch MB, et al. Language and trans health. *Int J Transgen.* (2017) 18:1–6. doi: 10.1080/15532739.2016.1262127
74. Intons-Peterson MJ, Reddel M. What do people ask about a neonate? *Dev Psychol.* (1984) 20:358–9. doi: 10.1037/0012-1649.20.3.358
75. Bradley SJ, Steiner BW, Zucker K, Doering RW, Sullivan J, Finegan JK, et al. Gender identity problems of children and adolescents: the establishment of a special clinic. *Can Psychiatr Assoc J.* (1978) 23:175–83. doi: 10.1177/070674377802300309
76. Zucker KJ. Comment on "Serving Transgender Youth: Challenges, Dilemmas, and Clinical Examples" by Tishelman et al. (2015). *Prof Psychol Res Pract.* (2015) 46:306. doi: 10.1037/pro0000030
77. Singh D. *A follow-up study of boys with gender identity disorder* (Unpublished doctoral dissertation), University of Toronto, Toronto, ON, Canada (2012).
78. Hollingshead AB. *Four Factor Index of Social Status*. Unpublished manuscript. New Haven, CT: Department of Sociology, Yale University (1975).
79. Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont Department of Psychiatry (1991).
80. Zucker KJ, Finegan JK, Doering RW, Bradley SJ. Human figure drawings of gender-problem children: a comparison to siblings, psychiatric, and normal controls. *J Abnorm Child Psychol.* (1983) 11:287–98. doi: 10.1007/BF00912092
81. Zucker KJ, Doering RW, Bradley SJ, Finegan JK. Sex-typed play in gender-disturbed children: a comparison to sibling and psychiatric controls. *Archiv Sexual Behav.* (1982) 11:309–21. doi: 10.1007/BF01541592
82. Alexander GM, Hines M. Gender labels and play styles: their relative contribution to children's selections of playmates. *Child Dev.* (1994) 65:869–79. doi: 10.2307/1131424
83. Fridell SR, Owen-Anderson A, Johnson LL, Bradley SJ, Zucker KJ. The playmate and play style preferences structured interview: a comparison of children with gender identity disorder and controls. *Archiv Sexual Behav.* (2006) 35:729–37. doi: 10.1007/s10508-006-9085-8
84. Zucker KJ, Lozinski JA, Bradley SJ, Doering RW. Sex-typed responses in the Rorschach protocols of children with gender identity disorder. *J Pers Assess.* (1992) 58:295–310. doi: 10.1207/s15327752jpa5802_9
85. Zucker KJ, Bradley SJ, Lowry Sullivan CB, Kuksis M, Birkenfeld-Adams A, Mitchell JN. A gender identity interview for children. *J Pers Assess.* (1993) 61:443–56. doi: 10.1207/s15327752jpa6103_2
86. Wallien MSC, Quilty LC, Steensma TD, Singh D, Lambert SL, Leroux A, et al. Cross-national replication of the gender identity interview for children. *J Pers Assess.* (2009) 91:545–52. doi: 10.1080/00223890903228463
87. Zucker KJ. Gender identity interview for children. In: Milhausen RR, Sakaluk JK, Fisher TD, Davis CM, Yarber WL, editors. *Handbook of Sexuality-Related Measures*, 4th ed. New York, NY: Routledge (2020). p. 325–8.
88. Cohen-Kettenis PT, Wallien M, Johnson LL, Owen-Anderson AFH, Bradley SJ, Zucker KJ. A parent-report gender identity questionnaire for children: a cross-national, cross-clinic comparative analysis. *Clin Child Psychol Psychiatry.* (2006) 7:433–56. doi: 10.1177/1359104506059135
89. Johnson LL, Bradley SJ, Birkenfeld-Adams AS, Kuksis MAR, Maing DM, Mitchell JN, et al. A parent-report gender identity questionnaire for children. *Archiv Sexual Behav.* (2004) 33:105–16. doi: 10.1023/B:ASEB.0000014325.68094.f3
90. Zucker KJ. Gender identity questionnaire for children. In: Milhausen RR, Sakaluk JK, Fisher TD, Davis CM, Yarber WL, editors. *Handbook of Sexuality-Related Measures*, 4th ed. New York, NY: Routledge (2020). p. 329–34.
91. Spencer D, Pasterski V, Neufeld S, Glover V, O'Connor TG, Hindmarsh PC, et al. Prenatal androgen exposure and children's

aggressive behavior and activity level. *Horm Behav.* (2007) 96:156–65. doi: 10.1016/j.yhbeh.2017.09.012
92. Zucker KJ, Wood H. Assessment of gender variance in children. *Child Adolesc Psychiatr Clin N Am.* (2011) 20:665–80. doi: 10.1016/j.chc.2011.07.006
93. Sattler JM. *Assessment of Children: Cognitive Applications*, 4th ed. San Diego, CA: Jerome M. Sattler Publisher, Inc. (2001).
94. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, DC: American Psychiatric Association (1994).
95. Deogracias JJ, Johnson LL, Meyer-Bahlburg HFL, Kessler SJ, Schober JM, Zucker KJ. The gender identity/gender dysphoria questionnaire for adolescents and adults. *J Sex Res.* (2007) 44:370–9. doi: 10.1080/00224490701586730
96. Singh D, Deogracias JJ, Johnson LL, Bradley SJ, Kibblewhite SJ, Meyer-Bahlburg HFL, et al. The gender identity/gender dysphoria questionnaire for adolescents and adults: further validity evidence. *J Sex Res.* (2010) 47:49–58. doi: 10.1080/00224490902898728
97. Zucker KJ, Meyer-Bahlburg HFL, Kessler SJ, Schober J. Gender identity/gender dysphoria questionnaire for adolescents and adults. In: Milhausen RR, Sakaluk JK, Fisher TD, Davis CM, Yarber WL, editors. *Handbook of Sexuality-Related Measures*, 4th ed. New York, NY: Routledge (2020). p. 343–50.
98. Zucker KJ, Bradley SJ, Oliver G, Blake J, Fleming S, Hood J. Psychosexual development of women with congenital adrenal hyperplasia. *Horm Behav.* (1996) 30:300–18. doi: 10.1006/hbeh.1996.0038
99. Schneider C, Cerwenka S, Nieder TO, Briken P, Cohen-Kettenis PT, De Cuypere G, et al. Measuring gender dysphoria: a multicenter examination and comparison of the Utrecht gender dysphoria scale and the gender identity/gender dysphoria questionnaire for adolescents and adults. *Archiv Sexual Behav.* (2016) 45:551–8. doi: 10.1007/s10508-016-0702-x
100. Siegmann EM, Müller T, Dziadeck I, Mühle C, Lenz B, Kornhuber J. Digit ratio (2D:4D) and transgender identity: new original data and a meta-analysis. *Sci Rep.* (2020) 10:19326. doi: 10.1038/s41598-020-72486-6
101. Zucker KJ, Bradley SJ, Owen-Anderson A, Kibblewhite SJ, Wood H, Singh D, et al. Demographics, behavior problems, and psychosexual characteristics of adolescents with gender identity disorder or transvestic fetishism. *J Sex Marital Ther.* (2012) 38:151–89. doi: 10.1080/0092623X.2011.611219
102. Zucker KJ, Bradley SJ, Owen-Anderson A, Singh D, Blanchard R, Bain J. Puberty-blocking hormonal therapy for adolescents with gender identity disorder: a descriptive clinical study. *J Gay Lesbian Mental Health.* (2011) 15:58–82. doi: 10.1080/19359705.2011.530574
103. Storms MD. Theories of sexual orientation. *J Pers Soc Psychol.* (1980) 38:783–92. doi: 10.1037/0022-3514.38.5.783
104. Kinsey AC, Pomeroy WB, Martin CE. *Sexual Behavior in the Human Male*. Philadelphia, PA: W. B. Saunders (1948).
105. Langevin R. *Sexual Strands: Understanding and Treating Sexual Anomalies in Men*. Hillsdale, NJ: Erlbaum (1985).
106. King MF, Brunner GC. Social desirability bias: a neglected aspect of validity testing. *Psychol Market.* (2000) 17:79–103. doi: 10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0
107. Crowne DP, Marlowe D. A new scale of social desirability independent of psychopathology. *J Consult Psychol.* (1960) 24:349–54. doi: 10.1037/h0047358
108. Strahan R, Gerbasi KC. Short, homogeneous versions of Marlowe-Crowne social desirability scale. *J Clin Psychol.* (1972) 28:191–3. doi: 10.1002/1097-4679(197204)28:2<191::AID-JCLP2270280220>3.0.CO;2-G
109. Holden RR, Fekken GC. Three common social desirability scales: friends, acquaintances, or strangers? *J Res Pers.* (1989) 23:180–91. doi: 10.1016/0092-6566(89)90022-6
110. Silverthorn NA, Gekoski WL. Social desirability effects on measures of adjustment to university, independence from parents, and self-efficacy. *J Clin Psychol.* (1995) 51:244–251. doi: 10.1002/1097-4679(199503)51:2<244::AID-JCLP2270510214>3.0.CO;2-Q
111. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally & Company (1969).
112. Coleman E, Bockting W, Botzer M, Cohen-Kettenis P, DeCuypere G, Feldman J, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgen.* (2011) 13:165–232. doi: 10.1080/15532739.2011.700873
113. Arcelus J, Bouman WP, Van Den Noortgate W, Claes L, Witcomb G, Fernandez-Aranda F. Systematic review and meta-analysis of prevalence studies in transsexualism. *Eur Psychiatry.* (2015) 30:807–15. doi: 10.1016/j.eurpsy.2015.04.005
114. Zhang Q, Goodman M, Adams N, Corneil T, Hashemid L, Kreukels B, et al. Epidemiological considerations in transgender health: a systematic review with focus on higher quality data. *Int J Transgen.* (2020) 21:125–37. doi: 10.1080/26895269.2020.1753136
115. DeMaris A. *Logit Modeling: Practical Applications*. Newbury Park, CA: Sage (1992).
116. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. *Nat Hum Behav.* (2018) 2:6–10. doi: 10.1038/s41562-017-0189-z
117. Steensma TD, Cohen-Kettenis PT. More than two developmental pathways in children with gender dysphoria? *J Am Acad Child Adolesc Psychiatry.* (2015) 54:147–8. doi: 10.1016/j.jaac.2014.10.016
118. Li G, Kung KTF, Hines M. Childhood gender-typed behavior and adolescent sexual orientation: a longitudinal population-based study. *Dev Psychol.* (2017) 53:764–77. doi: 10.1037/dev0000281
119. Steensma TD, van der Ende J, Verhulst FC, Cohen-Kettenis PT. Gender variance in childhood and sexual orientation in adulthood: a prospective study. *J Sex Med.* (2013) 10:2723–33. doi: 10.1111/j.1743-6109.2012.02701.x
120. Xu Y, Norton S, Rahman Q. Childhood maltreatment, gender nonconformity, and adolescent sexual orientation: a prospective birth cohort study. *Child Dev.* (2020) 91:e984–94. doi: 10.1111/cdev.13317
121. Nyberg KL, Alston JP. Analysis of public attitudes towards homosexuality. *J Homosex.* (1976) 2:99–107. doi: 10.1300/J082v02n02_01
122. Hellman RE, Green R, Gray JL, Williams K. Childhood sexual identity, childhood religiosity, and 'homophobia' as influences in the development of transsexualism, homosexuality, and heterosexuality. *Archiv Gen Psychiatry.* (1981) 38:910–5. doi: 10.1001/archpsyc.1981.01780330068007
123. Zucker KJ, Owen A, Bradley SJ, Ameeriar L. Gender-dysphoric children and adolescents: a comparative analysis of demographic characteristics and behavior problems. *Clin Child Psychol Psychiatry.* (2002) 7:398–411. doi: 10.1177/1359104502007003007
124. Green R, Newman LE, Stoller RJ. Treatment of boyhood "transsexualism." An interim report of four years' experience. *Archiv Gen Psychiatry.* (1972) 26:213–7. doi: 10.1001/archpsyc.1972.01750210021003
125. Meyer-Bahlburg HFL. Gender identity disorder in young boys: a parent- and peer-based treatment protocol. *Clin Child Psychol Psychiatry.* (2002) 7:360–76. doi: 10.1177/1359104502007003005
126. Newman LE. Treatment for the parents of feminine boys. *Am J Psychiatry.* (1976) 133:683–7. doi: 10.1176/ajp.133.6.683
127. Zucker KJ. Cross-gender-identified children. In: Steiner BW, editor. *Gender Dysphoria: Development, Research, Management*. New York, NY: Plenum Press (1985). p. 75–174. doi: 10.1007/978-1-4684-4784-2_4
128. Zucker KJ. Gender identity disorder in children and adolescents. In: Gabbard GO, editor. *Treatments of Psychiatric Disorders*, 3rd ed., Vol. 2. Washington, DC: American Psychiatric Press (2001). p. 2069–94.
129. Drescher J, Pula J. Ethical issues raised by the treatment of gender-variant prepubescent children. *Hast Center Rep.* (2014) 44:S17–22. doi: 10.1002/hast.365
130. Giordano S. *Children With Gender Identity Disorder: A Clinical, Ethical, and Legal Analysis*. New York, NY: Taylor & Francis Group (2013).
131. Pleak RR. Ethical issues in diagnosing and treating gender-dysphoric children and adolescents. In: Rottnek M, editor. *Sissies and Tomboys: Gender Nonconformity and Homosexual Childhood*. New York, NY: New York University Press (1999). p. 34–51.
132. Stein E. Commentary on the treatment of gender variant and gender dysphoric children and adolescents: common themes and ethical reflections. *J Homosex.* (2012) 59:480–500. doi: 10.1080/00918369.2012.653316
133. Zucker KJ. Treatment of gender identity disorders in children. In: Blanchard R, Steiner BW, editors. *Clinical Management of Gender Identity Disorders in Children and Adults*. Washington, DC: American Psychiatric Press (1990). p. 25–45.

134. Rekers GA, Lovaas OL. Behavioral treatment of deviant sex role behaviors in a male child. *J Appl Behav Anal.* (1974) 7:173–90. doi: 10.1901/jaba.1974.7-173
135. Rekers GA. *Shaping Your Child's Sexual Identity*. Grand Rapids, MI: Baker Book House (1982).
136. Zucker KJ, Wood H, Singh D, Bradley SJ. A developmental, biopsychosocial model for the treatment of children with gender identity disorder. *J Homosex.* (2012) 59:369–97. doi: 10.1080/00918369.2012.653309
137. American Academy of Child and Adolescent Psychiatry. Practice parameter on gay, lesbian, or bisexual sexual orientation, gender nonconformity, and gender discordance in children and adolescents. *J Am Acad Child Adolesc Psychiatry.* (2012) 51:957–74. doi: 10.1016/j.jaac.2012.07.004
138. American Psychological Association. Guidelines for psychological practice with transgender and gender nonconforming people. *Am Psychol.* (2015) 70:832–864. doi: 10.1037/a0039906
139. Byne W, Bradley SJ, Coleman E, Eyler AE, Green R, Menvielle EJ, et al. Report of the American Psychiatric Association Task Force on treatment of gender identity disorder. *Archiv Sexual Behav.* (2012) 41:759–96. doi: 10.1007/s10508-012-9975-x
140. Zucker KJ. Gender identity disorder. In: Weiner IB, editor. *Adult Psychopathology Case Studies*. New York, NY: Wiley (2004). p. 207–28.
141. Zucker KJ. "I'm half-boy, half-girl": play psychotherapy and parent counseling for gender identity disorder. In: Spitzer RL, First MB, Williams JBW, Gibbons M, editors. *DSM-IV-TR® Casebook, Volume 2. Experts Tell How They Treated Their Own Patients*. Washington, DC: American Psychiatric Publishing (2006). p. 321–34.
142. Cohen-Kettenis PT, van Goozen SHM. Pubertal delay as an aid in diagnosis and treatment of a transsexual adolescent. *Eur Child Adolesc Psychiatry.* (1998) 7:246–8. doi: 10.1007/s007870050073
143. Gooren L, Delemarre-van de Waal H. The feasibility of endocrine interventions in juvenile transsexuals. *J Psychol Hum Sex.* (1996) 8:69–84. doi: 10.1300/J056v08n04_05
144. Ehrensaft D. Found in transition: our littlest transgender people. *Contemp Psychoanal.* (2014) 50:571–92. doi: 10.1080/00107530.2014.942591
145. Ehrensaft D. Treatment paradigms for prepubertal children. In: Forcier M, Van Schalkwyk G, Turban JL, editors. *Pediatric Gender Identity: Gender-Affirming Care for Transgender & Gender Diverse Youth*. New York, NY: Springer (2020). p. 171–85. doi: 10.1007/978-3-030-38909-3_13
146. Wong W, Chang SCH. Social transitioning for gender dysphoric children. In: Lev AI, Gottlieb AR, editors. *Families in Transition: Parenting Gender Diverse Children, Adolescents, and Young Adults*. New York, NY: Harrington Park Press (2019). p. 356–73.
147. Wong WI, van der Miesen AIR, Li TGF, MacMullin LN, VanderLaan DP. Childhood social gender transition and psychosocial well-being: a comparison to cisgender gender-variant children. *Clin Pract Pediatr Psychol.* (2019) 7:241–53. doi: 10.1037/cpp0000295
148. Chen D, Edwards-Leeper L, Stancin T, Tishelman A. Advancing the practice of pediatric psychology with transgender youth: state of the science, ongoing controversies, and future directions. *Clin Pract Pediatr Psychol.* (2018) 6:73–83. doi: 10.1037/cpp0000229
149. Dreger A. Gender identity disorder in childhood: inconclusive advice to parents. *Hast Center Rep.* (2009) 39:26-9. doi: 10.1353/hcr.0.0102
150. Green R. To transition or not to transition? That is the question. *Curr Sexual Health Rep.* (2017) 9:79–83. doi: 10.1007/s11930-017-0106-5
151. Olson KR, Blotner C, Alonso D, Lewis K, Edwards D, Durwood L. Family discussions of early childhood social transitions. *Clin Pract Pediatr Psychol.* (2019) 7:22–40. doi: 10.1037/cpp0000289
152. Zucker KJ. Different strokes for different folks. *Child Adolesc Mental Health.* (2020) 25:36–7. doi: 10.1111/camh.12330
153. Rae JR, Gulgoz S, Durwood L, DeMeules M, Lowe R, Lindquist G, et al. Predicting early-childhood gender transitions. *Psychol Sci.* (2019) 30:669–81. doi: 10.1177/0956797619830649
154. Tollit MA, Pace CC, Telfer M, Hoq M, Bryson J, Fulkoski N, et al. What are the health outcomes of trans and gender diverse young people in Australia? Study protocol for the Trans20 longitudinal cohort study. *BMJ Open.* (2019) 9:e032151. doi: 10.1136/bmjopen-2019-032151

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer RB declared a past co-authorship with one of the authors KZ to the handling Editor.

Grootens-Wiegers et al. BMC Pediatrics (2017) 17:120
DOI 10.1186/s12887-017-0869-x

BMC Pediatrics

REVIEW Open Access

Medical decision-making in children and adolescents: developmental and neuroscientific aspects

Petronella Grootens-Wiegers[1,2*], Irma M. Hein[3], Jos M. van den Broek[1,2] and Martine C. de Vries[4,5]

Abstract

Background: Various international laws and guidelines stress the importance of respecting the developing autonomy of children and involving minors in decision-making regarding treatment and research participation. However, no universal agreement exists as to at what age minors should be deemed decision-making competent. Minors of the same age may show different levels of maturity. In addition, patients deemed rational conversation-partners as a child can suddenly become noncompliant as an adolescent. Age, context and development all play a role in decision-making competence. In this article we adopt a perspective on competence that specifically focuses on the impact of brain development on the child's decision-making process.

Main body: We believe that the discussion on decision-making competence of minors can greatly benefit from a multidisciplinary approach. We adopted such an approach in order to contribute to the understanding on how to deal with children in decision-making situations. Evidence emerging from neuroscience research concerning the developing brain structures in minors is combined with insights from various other fields, such as psychology, decision-making science and ethics. Four capacities have been described that are required for (medical) decision-making: (1) communicating a choice; (2) understanding; (3) reasoning; and (4) appreciation. Each capacity is related to a number of specific skills and abilities that need to be sufficiently developed to support the capacity. Based on this approach it can be concluded that at the age of 12 children can have the capacity to be decision-making competent. However, this age coincides with the onset of adolescence. Early development of the brain's reward system combined with late development of the control system diminishes decision-making competence in adolescents in specific contexts. We conclude that even adolescents possessing capacities required for decision-making, may need support of facilitating environmental factors.

Conclusion: This paper intends to offer insight in neuroscientific mechanisms underlying the medical decision-making capacities in minors and to stimulate practices for optimal involvement of minors. Developing minors become increasingly capable of decision-making, but the neurobiological development in adolescence affects competence in specific contexts. Adequate support should be offered in order to create a context in which minors can make competently make decisions.

Keywords: Decision-making, Neuroscience, Competence, Children, Adolescents, Brain development, Minors

* Correspondence: p.grootens@vu.nl; grootensp@gmail.com
[1]Science Communication and Society, Leiden University, Leiden, The Netherlands
[2]Athena Institute for Research on Innovation and Communication in Health and Life Sciences, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, The Netherlands
Full list of author information is available at the end of the article

Background

Various international guidelines stress the importance of involving children in decision-making regarding medical treatments and research participation. According to article 12 of the UN Convention on the Rights of the Child, "children shall be provided with the opportunity to be heard in any judicial or administrative proceeding affecting the child directly" [1]. More specific medical guidelines include The Second Directive by the European Parliament and the Council of the European Union, which states "A clinical trial on minors may be undertaken only if [] the minor has received information according to its capacity of understanding" [2]. In addition, many countries have laws specifying at what age children should be involved in decisions about medical treatment or scientific research. In the Netherlands for example, children from the age of 16 may take treatment decisions independently, and children from the age of 12 are allowed to give informed consent for research participation or treatment together with their parents. In the US a minimum age of 7 years old is defined for asking assent (as opposed to legal consent) from children [3]. In the UK, children under the age of 16 cannot be treated without parental consent, unless they prove to be mature according to the Gillick ruling [4].

These laws and guidelines underline the importance of respecting the developing autonomy of children. However, they also show that there is no universal agreement as to at what age it is appropriate for children to be considered competent for decision-making. Empirical evidence demonstrates that children have an emerging competence at a very young age. Weithorn & Campbell found children as young as 9 years old to have the capacity to make informed choices [5]. In addition, some studies conclude that children at age 14 or 15 are as competent as adults [5–7]. A recent study demonstrated that generally children older than 11.2 years may be competent to consent to clinical research [8]. Yet in most countries, children are considered incompetent until the age of 18 or 21, when they officially have reached legal adulthood.

In medical practice it is not clear-cut whether a child of a certain age is sufficiently competent for medical decision-making. Different children of the same age may have a different level of maturity. Young children, who have demonstrated sufficient competence for decision-making in a certain situation, can lack adequate competence in another. Furthermore, children who have shown to be reasonable conversation-partners during their treatment, can (temporarily) be noncompliant in adolescence, as illustrated by the story of Elsa in Table 1. Therefore, in this article we explore a way in which insights in brain development can contribute to insights in decision-making competence of children at various ages.

Table 1 The story of Elsa

Elsa is a 16 year old adolescent who was diagnosed with diabetes type I at the age of 4.

The first years after the diagnosis, Elsa's parents did all the diabetes care. They measured blood sugars, and adjusted insulin dose as required during meals, exercise etc. The insulin pump Elsa was wearing had a child safety lock to prevent accidental use by Elsa. Elsa was able to express how she felt about the disease but did not have any influence in the treatment. When Elsa became older, she was very eager to learn about her daily diabetes care. From the age of 7 she was taught how to measure her own blood sugar and what the result meant. From about the age of 8, she could instruct the pump to give the insulin dose needed during meals (as long as her parents had written down in her lunch box how many carbohydrates were in the lunch). At 10, Elsa showed profound insight in how to adjust her insulin pump settings when her blood glucose levels were not optimal. By then, she was so well informed and experienced that she was able to handle her diabetes with her parents only exerting global supervision.

When Elsa turned 12 and went to secondary school things changed. She started to exert less self-control. She did not measure glucose levels and did not inject insulin for meals at school. Her school friends were unaware of her diabetes because Elsa did not inform them. Elsa tried to deny her diabetes at school, and often even took off the insulin pump, for example during physical exercise at school.

When at the pediatrician's office, Elsa was always friendly, showing remorse and promising improvement. At age 14 however, she had to be admitted to the Intensive Care Unit because of severe dysregulation of her diabetes and an acute life-threatening situation. At age 16, the same happened after drinking large amounts of alcohol

Decision-making competence and capacity

A certain level of competence is required for medical decision-making in order to balance the respect for autonomy with the protection of vulnerable patients [9]. In order to be sufficiently *competent*, one needs to have the *mental capacity* to make decisions, but also should be accountable of the decision in the specific situation. That is, one can in theory have the mental ability to make a reasonable decision, but a certain situation can reduce a person's competence, e.g. due to stress or peer pressure [10]. *Decision-making capacity* is thus necessary, but not sufficient for being *decision-making competent*.

Decision-making capacity can be defined by four standards: (1) expressing a choice; (2) understanding; (3) reasoning; and (4) appreciation [11–13]. In order to be considered competent to make a decision all four capacity standards should be met [11, 13].

However, decision-making competence is not an on-or-off phenomenon [14], but is relative to the specific decision in the specific situation [14, 15]. Furthermore, certain diseases, medical as well as mental, can affect competence, either temporarily (e.g. when a patient loses consciousness) or in a chronical manner (as is the case in progressing Alzheimer's disease [16]).

Miller has proposed a model on children's capacity in which initial predisposing factors are identified, followed by four groups of factors that influence decision-making

competence, namely child, parent, clinician, and situational factors [10]. Predisposing factors include the discussed cognitive development, as well as experience. Factors related to the child are personality [17], and emotional state of the child that can affect capacity and serve as a spotlight or motivator for information and preferences [6, 17, 18]. In addition, disease severity can affect understanding, as well as retention of information and reasons to consent [19]. Parent and clinicians can influence the child's competence with their attitude towards the child and the attention and support provided in the decision-making process [6, 17, 20]. Finally, situational factors as the type and complexity of the decision, the setting and time constraints play a role [10].

In Miller's model, (cognitive) development is thus an important predisposing factor for decision-making competence in children. As children grow older, their capacities to comprehend information and therefore competence to make a decision increase. Therefore, insight in the development of various abilities related to medical decision-making may contribute to understanding at what age children could be considered decision-making competent.

Aim

We believe that the discussion about decision-making competence of minors can greatly benefit from a multidisciplinary approach, as the issue has many aspects. We reviewed the evidence emerging from neuroscience research concerning the impact of developing brain structures on children's decision-making capacities and competence. We subsequently combined insights from neuroscience with various other fields: psychology, decision-making science, ethics and medical practice. It is not our aim to quantify specifically at what age exactly children should be considered decision-making competent, but rather to contribute to insights on how to deal with children in medical decision-making, and to add to the general discussion on children and decision-making.

In this paper, we will discuss the aforementioned four standards of medical decision-making capacity as defined by Appelbaum and Grisso [13]. We will discuss the development of the various skills and abilities that are required for each standard according to Appelbaum and Grisso [13], as well as describe the brain areas that are involved in these skills. Relating brain areas, development and decision-making abilities can contribute to an understanding of child behavior and competence. However, we will only be able to provide a simplified insight in the neuroscience background, as each ability requires the contribution of numerous brain areas and structures and we aim to keep this discussion readable for clinicians without a background in neuroscience. For a more elaborate overview of brain structures involved in decision-making, we want to point the reader to the paper of Rosenbloom et al. [21].

In addition, we will discuss what happens in the brain during adolescence and how this influences decision-making. Adolescents often seem to have a reduced ability to make reasonable decisions [22, 23], and this phenomenon can be related to the developmental events happening in the brain during this period. The paragraph on adolescents will enlighten why many adolescent patients will consent to treatment in the clinic but do not do as asked when they return to normal day-to-day life, such as in the story of Elsa (Table 1).

Development of abilities and brain areas related to the four capacity standards

The four standards of medical decision-making capacity will be discussed in association with neurological skills. In this section, the main course of development is discussed, a more detailed discussion of the neurological skills and related brain areas is provided in the Appendix (Additional file 1).

Expressing a choice

The first and least rigorous standard for decision-making capacity is the ability to express a choice. This standard implies that someone can communicate a preference of treatment or research participation, which is legally restricted to spoken or written language. The required neurological skill for this standard is being able to **communicate**, either in spoken language or nonverbally [13, 24]. Nonverbal communication can be used as an indication of dissent or of implicit consent, but not as a legal form of consent. Therefore this capacity is mainly related to verbal language development, which initiates in early childhood. From the age of 5, children have reasonable understanding of language, with refinement thereof continuing to the age of 9 and further throughout adolescence [25].

Understanding

The second standard requires the ability to understand the information provided about the proposed medical treatment or research and comprehending the fact that a choice needs to be made. Understanding requires a combination of neurological skills [13, 24]: One first needs to have sufficient **intelligence** and **language** proficiency to process the information. Further, one needs to be able to **orient** and direct **attention** towards the information. In addition, understanding requires **memory** and **recall** skills, in order to process and integrate information beyond the short-term moment. The foundation for these skills is laid down in the first year of life. Maturity in orienting and attention develops around the ages of 7–10 [26–28]. During childhood the ability to remember information and the amount that can be remembered develops. Memory specifically increases between the ages of 6 and 12, and then goes on to

slightly increase during adolescence [29, 30]. Children at the age of 10–12 appear to have recall abilities compared to adults [31–33].

Reasoning

The third standard is that, next to understanding the factual information, someone should be able to reason about risks, benefits and possible consequences of the treatment or research options presented [11, 13, 24]. This standard is a step further from factual understanding and requires the ability for logical **reasoning** and **weighing risks and benefits**. Children at the age of 6 to 8 already demonstrate the ability for logic reasoning [34, 35]. Between the ages of 8 and 11, children's reasoning skills improve significantly, mainly due to improved use and access to their own knowledge [36]. Complex reasoning about alternative causal relations needs more time to develop, in adolescence is has become more accurate, but even adults often make mistakes [34]. Risk identification develops strongly between the ages of 6 and 10 [37]. Adults are better in identifying risks than children and adolescents but not in identification of benefits [38]. In addition, as will be discussed later in this paper, even though risk identification is mature in late adolescence, the way people of this age will deal with risks differs from that of adults.

Appreciation

The strictest standard of decision-making capacity is appreciation. The appreciation of the nature of a situation implies that someone will not only understand the various options, but also the relevance of these options for the personal situation. In order to appreciate the situation and personal relevance of the decision at hand, one needs to have the ability of **abstract thinking**, which includes being aware that others have a mind of their own, which is called **theory of mind** [12, 13]. Abstract thinking, about things that are intangible, is necessary to understand the consequences of a decision. There are many different skills and brain areas involved in this skill. Between the age of 3 and 4, children already start to recognize their own beliefs and desires, which contribute to the development of personal norms and values, and start to understand how these influence their actions [25, 39, 40]. Improvement of the efficiency of working memory with age further increases the ability to think about abstract and hypothetical things, situations and norms and values [34, 41].

Model

Below the discussed abilities and their developmental trajectories are visualized in a model (see Fig. 1). This overview shows that the necessary abilities and relating brain areas do not develop synchronically; some aspects of capacity are mature much earlier than others. This illustrates that decision-making competence is not an on-or-off concept, but rather a growing skill with age.

Translation of the model into clinical practice

The four capacity standards can be measured in clinical practice with the use of the MacArthur Competence Assessment Tool (MacCAT) [13], which is validated and used among adults. Results of a recent study on children's competence to consent to clinical research showed that the MacCAT could also be validly and reliably used in children. The MacCAT-CR was studied in a population of pediatric patients between 6 and 18 years of age [8]. The study demonstrated that age limits for children to be deemed competent to decide on research participation could be estimated as follows: children of 11.2 years and above generally appeared to be competent, whereas children of 9.6 years and younger were generally not competent. A change-over occurred between 9.6 and 11.2 years, and the cross-over point was estimated at 10.4 years [8]. In the same study, the four domains representing competence in most jurisdictions (understanding, appreciation, reasoning, and expressing a choice) appeared to constitute a single trait in children. These results correspond well with the model in Fig. 1. Below the age of 10, too many abilities are still in their (early) development and overall competence cannot be expected. However, as discussed, the cut-off age of 11.2 does not automatically imply competence for any decision in any situation. Rather, this age serves as an indication at what age competence might be expected given favorable environmental factors. In addition, an important influence on competence is the rise of adolescence, which is accompanied with very specific events in brain development.

Adolescence and decision-making competence

The demonstrated model might suggest a linear pattern in development and a corresponding linear increase in decision-making competence with age. However, due to differences in cross-talk between the various brain structures over the course of brain development, competence might fluctuate. A period in which this is especially pronounced is adolescence. In this period, great changes and developmental leaps take place in the brain, which can have a profound effect on decision-making competence.

Adolescence is a period associated with a number of health issues and increased mortality [42, 43]. Adolescents often have increased appetite and therefore a change in diet; in addition adolescence is typically the time where tobacco addiction initiates and a time of emerging alcohol and substance (ab)use [42, 44]. Further, for chronically ill children, this is a time where the disease management approach can change, sometimes creating risky of even

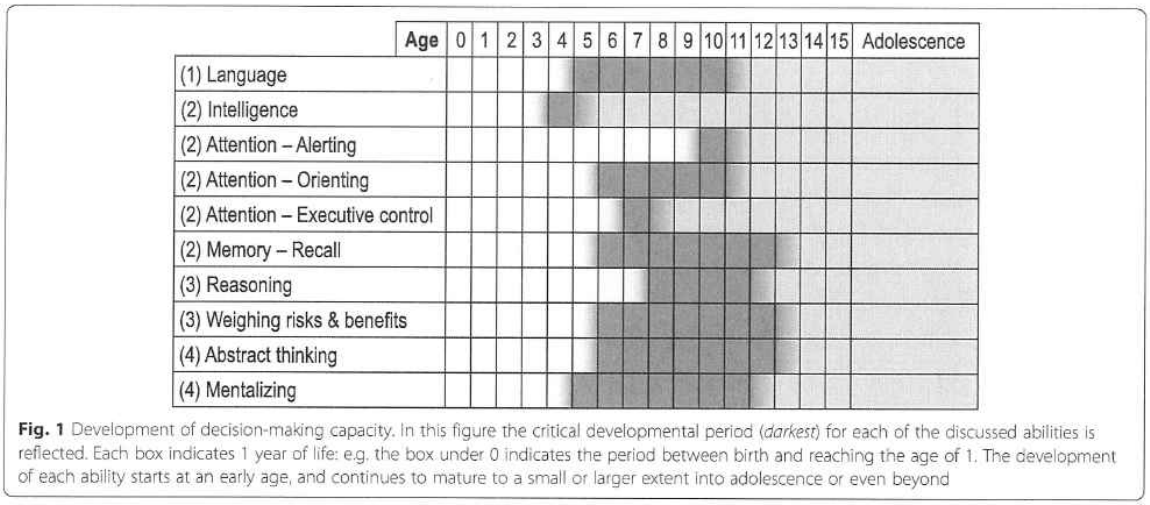


Fig. 1 Development of decision-making capacity. In this figure the critical developmental period (*darkest*) for each of the discussed abilities is reflected. Each box indicates 1 year of life: e.g. the box under 0 indicates the period between birth and reaching the age of 1. The development of each ability starts at an early age, and continues to mature to a small or larger extent into adolescence or even beyond

life-threatening situations, as illustrated in Table 1. The increased mortality seen in adolescence is mostly associated with risky behavior, sensation-seeking and peer influences affecting decision-making [43].

Adolescence starts around the age of 12 and the neurologic developments initiated can continue into early adulthood [7, 45]. The brain in adolescence differs significantly from the brain in childhood and adulthood [45–48]. To gain more insight in the effect of adolescence on decision-making, it is important to have an understanding of this period. The most significant changes in the brain are associated with processing rewards and risks, self-regulation, and the effect of peers on decision-making. These neurologic changes affect decision-making in general and, depending upon context, can affect medical decision-making to a certain extent as well.

Risk, sensation-seeking and self-regulation

Adolescents are prone towards increased risk-taking and this is associated with the development of a number of brain-structures. Two brain systems are especially important: the prefrontal cortex (PFC), which is the control system; and the ventral striatum, the reward system. The control system is involved in impulse control, the ability to stop a certain urge or action, and thus involved in self-regulation. The ability for self-regulation develops strongly from the age of 12 until the age of 18 [45], but continues to improve into early adulthood [7]. In addition, the prefrontal cortex also performs better at other functions that require control, such as planning ahead, weighing risks and benefits and in processing complicated decisions. The cross-talk between the control system and the reward system and associated emotional regulation is not fully developed before early adulthood [7]. This means that even though an adolescent can have intellectual maturity, this does not automatically imply the presence of emotional and social maturity [7, 47].

The reward system involves a structure that creates dopamine in response to rewards. Dopamine gives a feeling of pleasure, which can lead to learning and the urge to repeat the experience. During adolescence, the reward system becomes hyperresponsive, the dopamine response to a reward is much higher [49]. This is associated with increased reward-seeking and sensation-seeking [48–50]. The increased responsiveness of the reward system even applies to small rewards, making the positive effect of a small 'success' of a decision more pronounced for adolescents than for children or adults [49]. Thus in a dilemma in which there is a small chance of a reward, this reward can be attributed such a high value that the situation is no longer perceived as a dilemma by the adolescent and there is only one path to choose [22].

The development of the control and the reward systems do not follow a linear pattern, The last brain areas to mature are those involved in executive function and attention, located in the PFC [51]. Based on structural brain development research, there appears to be a 'mismatch' between the development of various regions, specifically the amygdala and the PFC. The amygdala, responsible for emotion processing and input in the reward system, starts to mature in late childhood and stabilizes at mid- to late adolescence [52]. However, the PFC starts to mature in early adolescence and it is not until young adulthood that this area is mature. In addition, the nucleus accumbens in the ventral striatum, appears to develop early in some and later in others, which might explain a 'mismatch' in some adolescents [52].

Thus, the control system (PFC) develops slowly, even into early adulthood whereas the reward system (amygdala and possibly nucleus accumbens) already changes in early adolescence [7]. This nonlinear development accounts for the risky decisions often observed in adolescents, such as binge drinking or drunk driving [22]. This is not to say that adolescents are incapable of estimating risks or making responsible decisions. Evidence from laboratory experiments demonstrates that adolescents have a decision-making capacity similar to adults [7, 44, 47, 53]. Adolescents thus have better insight in decision-making than children do, consistent with our proposed model. Yet do they end up in precarious and risky situations and their behavior is often not consistent with their capacities.

This inconsistency can be explained with the distinction between 'hot' and 'cold' contexts. An emotional context is called a 'hot' situation, whereas in 'cold' situations, decisions are not or only minimally emotionally loaded [22]. When emotions play a role in a situation, this can significantly influence the decision-making process and outcome [44, 54]. Whether a situation is hot or cold is not predefined: it can vary per individual to what extent a context is perceived as emotionally loaded [48]. Research has shown that during adolescence, risk-taking in decisions in cold situations is similar to that of children and adults [48]. However, when in a hot situation, risk-taking is increased, affecting decision-making severely [7, 44, 48]. This explains the often risky decisions that adolescents make, seemingly only thinking about short-term rewards, even though afterwards they can reasonably assess their leap in judgment.

One particular type of emotionally loaded situation is the presence of peers. As adolescence is essentially a process to develop the capacity to navigate the social landscape, social cues become increasingly important [53]. During adolescence, the acceptation by peers becomes an important purpose in everyday life and guides decision-making [55]. Correspondingly, the ability to understand the perspective of another person and predict that person's behavior increases [48]. As discussed, this ability for mentalizing develops until late adolescence and it modulates decision-making. In addition, self-awareness increases during adolescence [55].

Accordingly, decision-making in the presence of peers is substantially different from individual decisions [56]. When with peers, the brain sensitizes even more towards rewards and possible rewarding outcomes are higher valued. The adolescent can show an adequate understanding of the situation and its risks involved, but the developing control system can become overruled by the emotional cues in this 'hot' context [47]. As a result of the hot context adolescents are more prone towards making risky decisions, even when only a small reward can be expected [43]. This also explains why adolescents' risk-prone tendencies are mostly observed in group situations, especially when there is a certain form of excitement present ('hot') [22].

Strengths and vulnerabilities of adolescents in medical care

The developing brain in adolescence thus leads to lower cognitive control and leaves adolescents more prone towards risk-taking, especially when together with peers. These characteristics can affect decision-making competence during adolescence. The competence of adolescents to make a decision can vary per situation. Some medical decisions can be considered 'cold', with minimal influence of social or emotional factors [7], providing a good context for a competent decision. Treatment and research decisions are generally not impulsive decisions, and a certain amount of time for consideration is provided. This will reduce impulsive and unreasoned decisions in adolescents [47]. However, this does not mean that an adolescent will necessarily live up to the decision in the long run, as context might change. For example, a diabetes patient can be very aware of the benefits of a regular and structured diet and discuss this wisely in a hospital setting. However, living up to the treatment pattern can be much harder when the same person is with a group of friends who decide to skip class and go for a snack. Now the context of the decision turned into a hot, peer-influenced and exciting situation, which affects the decision-making rationale and possibly the outcome, as also illustrated in the example in Table 1. Some adolescents are more susceptible to such an effect than others, and thus the outcome of the dilemma is not necessarily the same for each young patient, making practice very unpredictable.

Especially in treatment situations, adolescents can demonstrate this type of seemingly decreased competence for responsible decisions [42]. Short-term rewards become more important than long-term rewards, even when choosing for an immediate reward can mean a loss on the long-term [48, 53]. This can make it complicated to stick to a healthy lifestyle or treatment pattern, which usually does not deliver immediate rewards, but is meant to increase long-term health. Another factor playing a role might be the expectation of the long-term reward. It appears that adolescents over-estimate their risk of dying soon [57]. This over-estimation of a chance on a short life automatically diminishes the value of any long-term rewards, as the chance of living long enough to receive the reward is considered relatively low.

Although these characteristics render adolescents more vulnerable towards risky situations and their consequences, they also are an important aspect of developing into an adult. During adolescence, the brain

shows a high amount of plasticity, resulting in vulnerabilities, but also in opportunities [43]. The sensitivity to rewards together with increased value of social cues creates a perfect situation for learning new skills that are important to function in a social context [53, 55]. Adolescents can learn very quickly and can sometimes even outperform adults when it comes to problem-solving and creativity [53]. In addition, adolescence is a time in which health behavior can be stimulated to consolidate, or when behavior can easily be altered, if the adolescent is motivated to do so [42]. Therefore, adolescence offers an opportunity to target health behavior and disease management and teach the brain new behavior [42, 58].

Discussion

In this paper we have addressed the complexity of assessing competence in minors and analyzed the neurological development of decision-making capacities based on the four standards from Appelbaum et al.; expressing a choice, understanding, reasoning, and appreciation [13]. The development of the brain demonstrates a non-linear pattern and therefore decision-making competence does not increase in a linear fashion with age. Based on our model, it might be expected that children around the age of 12 may already have the competence to make medical decisions. However, this age coincides with the onset of adolescence, which is associated with altered decision-making patterns. Adolescents are prone towards increased risk-taking, especially in emotional situations and when with peers. This affects their decision-making competence, mostly in 'hot' or emotional situations, such as compliance decisions in everyday life at school, but less so in 'cold' situations such as deciding upon treatment in the hospital. As a result, decision-making competence in adolescence can vary greatly between moments and contexts, as was illustrated by the story of Elsa (Table 1). It is thus complicated to pinpoint a certain age at which a child should be considered fully competent to make medical-decisions based on brain development. Even more so since brain development can vary between individuals and gender.

In addition, in this paper we mainly discuss the neurological background of decision-making competence, with the aim to contribute to insights about the age at which the brain is mature enough to be capable of making a decision. However, mature neurological capacity does not automatically mean that a child is competent for any medical decision. There are many factors that influence decision-making competence, either temporarily or chronically, as illustrated by the model of Miller on children's capacity, describing the predisposing factor of cognitive development, and in addition groups of factors revolving around the child itself, its parents, the clinican, and situational factors. It thus appears impossible to define a cut-off point at what age all children should be presumed competent to make medical decisions based on neuroscience. Nevertheless, based on empirical research, indications for a just age limit for alleged competence to consent in children were estimated. In the clinical research context, children of 11.2 years and above were generally competent. In the treatment context initial indications point into the direction of comparable age limits for alleged competence, around the age of 12, but more research is needed to confirm these findings.

The confirmed potential for competence, combined with the influence of other factors affecting competence, led to the recommendation of a double consent procedure (child and parent) for minors from the age of 12 until 18. Taking into account that parents are generally provided with the legal authority to raise their children, they are assigned with rights and responsibilities. A double consent procedure could achieve an equable consideration between the legal position of the child and that of the parents. A double consent procedure will do justice to both developmental aspects of children and the specific characteristics of the parent–child dyad. The parental role offers extra protection by creating the context for the child's competent decision-making and by facilitating the child's long term autonomy. In general, the perspective and attitudes of the adults (both parents and clinician) towards the child may be an important predisposing factor in order to stimulate the highest competence in the child [59]. How adults in the current social climate view minors, can affect whether they live up to their potential. Often children are considered merely on their way to adulthood, but not yet there, This might imply that they are 'less' than an adult and are incapable of understanding or forming opinions, let alone making decisions. When children are viewed this way, they will not be informed adequately and will not be supported optimally to take a role in decision-making that does justice to their potential. In order for children to be optimally comptetent, it is important for the involved adults to be aware that children have their own characteristics and perspectives, that are as valuable as (but not necessarily similar to) those of adults, and that they are informed and supported accordingly. This issue will be addressed in more depth in an upcoming paper by the authors (manuscript in preparation).

In accordance with the development of decision-making capacity, and out of respect for children's autonomy, children should be increasingly informed and involved in the decision-making process [5, 60, 61]. Attention should be paid to providing the child with adequate information, as decision-making competence is *'only as good as the provided information'* [14, 20]. This means that the information supplied needs to be adapted to the child's level of

communication and understanding, for example by providing separate sheets for the child and offering oral explanations [60, 62]. As long as there is no adequate information, it is certain that children cannot meaningfully be involved in the decision-making process [9, 59, 63–65].

Conclusion

Currently, medical laws and regulations reflect the belief that child development influences children's decision-making processes to the extent that age limits are presented at which children are deemed incompetent or competent. In problematic cases, child psychiatrists and –psychologists are consulted to assess the decision-making capacities of a child, the clinical operationalization of the legal concept of competence. In this article we adopt a perspective on such competence assessment that specifically focuses on the impact of brain development on the child's decision-making process. Taking this perspective opens up the opportunity to implement results from an emerging field in neurobiological research on how developing brain structures may affect a child's decision-making capacities. The insights provided in this paper are intended to aid insight in the practice of dealing with minors in medical situations, and to stimulate further discussion about decision-making capacity and competence in children.

In neuroscience, changes in brain structures have been detected that are related to changes in decision-making capacities. The authors are aware that this is a rapidly developing field, that is currently just starting to gain knowledge about the specific development of these abilities in the brain, with many questions left to be answered [18]. As neuroscience is a relatively new and developing science, this paper only provides initial insight in the issue, but evolving neuroscience will lead to further insights.

Summary

Various international laws and guidelines underline the importance of respecting the developing autonomy of children. However, they also show there is no universal agreement as to at what age children are considered competent for decision-making. In this article we adopt a perspective on competence that specifically focuses on the impact of brain development on the child's decision-making abilities. Neuroscience research is related to the 4 capacities required for medical decision-making, which are communicating a choice, understanding, reasoning, and appreciation. Based on this approach it can be concluded that at the age of 12 children may have the capacity to be decision-making competent, given favorable environmental factors. However, this age coincides with the onset of adolescence. Early development of the brain's reward system combined with late development of the control system diminishes decision-making competence in adolescents in specific contexts. We conclude that even adolescents possessing capacities required for decision-making, may need support of facilitating environmental factors.

Additional file

Additional file 1: Appendix [66–88]. (PDF 692 kb)

Abbreviations

ACC: Anterior Cingulate Cortex; DLPFC: Dorsal Lateral Prefrontal Cortex; LC: Locus Coeruleus; OFC: Orbitofrontal Cortex; PFC: Prefrontal Cortex

Acknowledgements

The authors kindly thank Barbara. R. Braams from the department of Developmental and Educational Psychology at Leiden University for her aid in preparing this paper. The authors would also like to thank Karlijn Groenendijk for her input in the first stages of the concept of this paper.

Funding

This study was funded by The Netherlands Organisation for Health Research and Development [project number 113203016].

Availability of data and materials

Not Applicable

Authors' contributions

PGW designed the concept of this review and drafted the manuscript. IMH provided her expertise to support the scientific soundness of the manuscript and contributed with extensive feedback on the line of reasoning in the manuscript. JMvdB was involved in the draft concept and revisions of the manuscript. MCdV was involved in designing the concept of this manuscript and provided extensive feedback on the line of reasoning of the draft versions. All authors read and approved the final manuscript.

Authors' information

PGW, PhD, is a research associate in health communication and ethics. She has developed and studied informed consent material for children and published her thesis "Targeted Informed Consent – Empowering young participants in medical-scientific research" on this topic at Leiden University.
IMH, PhD, is a child and adolescent psychiatrist, as well as researcher in the area of decision-making competence. She has developed a tool to assess children's competence to consent to medical treatment and scientific research.
JMvdB, PhD, is a professor in biomedical science communication at Leiden University, with a keen interest in visual communication and health communication for children as well as low-literate individuals.
MCdV, MD, PhD, is pediatrician and medical ethicist at the Leiden University Medical Centre. Her research interests include research ethics and child participation in decision making. She chairs the Committee on Ethics and Health Law of the Dutch Paediatric Association.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not Applicable.

Ethics approval and consent to participate

Not Applicable.

Publisher's Note

Author details
[1]Science Communication and Society, Leiden University, Leiden, The Netherlands. [2]Athena Institute for Research on Innovation and Communication in Health and Life Sciences, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, The Netherlands. [3]Child and Adolescent Psychiatry and de Bascule, Academic Medical Center Amsterdam, Amsterdam, The Netherlands. [4]Department of Medical Ethics and Health Law, Leiden University Medical Center, Leiden, The Netherlands. [5]Department of Pediatrics, Leiden University Medical Center, Leiden, The Netherlands.

Received: 6 August 2015 Accepted: 22 April 2017
Published online: 08 May 2017

References

1. Unicef: Convention on the Rights of the Child. United Nations, Treaty Series 1577.3 1989.
2. EU: Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the member states relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. In Official Journal of the European Communities. 2001:34–44.
3. National Institutes of Health: Children's assent to clinical trial participation. 2005. https://www.cancer.gov/about-cancer/treatment/clinical-trials/patient-safety/childrens-assent. Accessed 2 May 2017.
4. Mayer RE, Moreno R. Nine ways to reduce cognitive load in multimedia learning. Educ Psychol. 2003;38:43–52.
5. Weithorn LA, Campbell SB. The competency of children and adolescents to make informed treatment decisions. Child Dev. 1982;53:1589–98.
6. Mann L, Harmoni R, Power C. Adolescent decision-making: the development of competence. J Adolesc. 1989;12:265–78.
7. Steinberg L. Does recent research on adolescent brain development inform the mature minor doctrine? J Med Philos. 2013;38:256–67.
8. Hein IM, Troost PW, Lindeboom R. Accuracy of MacArthur Competence Assessment Tool for measuring children's competence to consent to clinical research. JAMA Pediatr. 2014;12(168):1147–53.
9. Appelbaum PS. Clinical practice. Assessment of patients' competence to consent to treatment. N Engl J Med. 2007;357:1834–40.
10. Miller VA, Drotar D, Kodish E. Children's competence for assent and consent: a review of empirical findings. Ethics Behav. 2004;14:255–95.
11. Grisso T, Appelbaum PS, Hill-Fotouhi C. The MacCAT-T: a clinical tool to assess patients' capacities to make treatment decisions. Psychiatr Serv. 1997; 48:1415–9.
12. Appelbaum PS, Roth LH. Competency to consent to research: a psychiatric overview. Arch Gen Psychiatry. 1982;39:951–8.
13. Appelbaum PS, Grisso T. The MacArthur Competence Assessment Tool for Clinical Research (MacCAT-CR). Sarasota: Professional Resource Press; 2001.
14. Ganzini L, Volicer L, Nelson WA, Fox E, Derse AR. Ten myths about decision-making capacity. J Am Med Dir Assoc. 2004;5:263–7.
15. Bolt IL, van Summeren MJ. Competence assessment in minors, illustrated by the case of bariatric surgery for morbidly obese children. Best Pract Res Clin Gastroenterol. 2014;28:293–302.
16. Marson DC, Ingram KK, Cody HA, Harrell LE. Assessing the competency of patients with Alzheimer's disease under different legal standards. A prototype instrument. Arch Neurol. 1995;52:949–54.
17. Alderson P. In the genes or in the stars? Children's competence to consent. J Med Ethics. 1992;18:119–24.
18. Weber EU, Johnson EJ. Mindful judgment and decision making. Annu Rev Psychol. 2009;60:53–85.
19. Schaeffer MH, Krantz DS, Wichman A, Masur H, Reed E, Vinicky JK. The impact of disease severity on the informed consent process in clinical research. Am J Med. 1996;100:261–8.
20. Hein IM, Troost PW, Lindeboom R, de Vries MC, Zwaan CM, Lindauer RJ. Assessing children's competence to consent in research by a standardized tool: a validity study. BMC Pediatr. 2012;12:156.
21. Rosenbloom MH, Schmahmann JD, Price BH. The functional neuroanatomy of decision-making. J Neuropsychiatry Clin Neurosci. 2012;24:266–77.
22. Steinberg L. Risk taking in adolescence: what changes, and why? Ann N Y Acad Sci. 2004;1021:51–8.
23. Dinwiddie R, Muller WG. Adolescent treatment compliance in asthma. J R Soc Med. 2002;95:68–71.
24. Appelbaum PS, Grisso T. Assessing patients' capacities to consent to treatment. N Engl J Med. 1988;319:1635–8.
25. Shaffer D, Kipp K. Developmental psychology. Belmont: Thomson Wadsworth; 2007.
26. Reed J, Warner-Rogers J, editors. Child Neuropsychology; Concept, Theory and Practice. Oxford: Wiley-Blackwell; 2008.
27. Rueda MR, Fan J, McCandliss BD, Halparin JD, Gruber DB, Lercari LP, Posner MI. Development of attentional networks in childhood. Neuropsychologia. 2004;42:1029–40.
28. Waszak F, Li SC, Hommel B. The development of attentional networks: cross-sectional findings from a life span sample. Dev Psychol. 2010;46:337–49.
29. Guillery-Girard B, Martins S, Deshayes S, Hertz-Pannier L, Chiron C, Jambaque I, Landeau B, Clochon P, Chetelat G, Eustache F. Developmental trajectories of associative memory from childhood to adulthood: a behavioral and neuroimaging study. Front Behav Neurosci. 2013;7:126.
30. Thaler NS, Goldstein G, Pettegrew JW, Luther JF, Reynolds CR, Allen DN. Developmental aspects of working and associative memory. Arch Clin Neuropsychol. 2013;28:348–55.
31. Rhodes SM, Murphy D, Hancock PJ. Developmental changes in the engagement of episodic retrieval processes and their relationship with working memory during the period of middle childhood. Br J Dev Psychol. 2011;29:865–82.
32. Sprondel V, Kipp KH, Mecklinger A. Developmental changes in item and source memory: evidence from an ERP recognition memory study with children, adolescents, and adults. Child Dev. 2011;82:1638–953.
33. Czernochowski D, Mecklinger A, Johansson M. Age-related changes in the control of episodic retrieval: an ERP study of recognition memory in children and adults. Dev Sci. 2009;12:1026–40.
34. Markovits H. The development of abstract conditional reasoning. In: Barrouillet P, Gauffroy C, editors. The development of thinking and reasoning. London: Psychology Press; 2013. p. 71–94.
35. Pillow BH, Pearson RM, Hecht M, Bremer A. Children's and adults' judgments of the certainty of deductive inferences, inductive inferences, and guesses. J Genet Psychol. 2010;171:203–17.
36. Markovits H, Fleury ML, Quinn S, Venet M. The development of conditional reasoning and the structure of semantic memory. Child Dev. 1998;69:742–55.
37. Hillier LM, Morrongiello BA. Age and gender differences in school-age children's appraisals of injury risk. J Pediatr Psychol. 1998;23:229–38.
38. Halpern-Felscher BL, Cauffman E. Costs and benefits of a decision. Decision-making competence in adolescents and adults. Appl Dev Psychol. 2001;22: 257–73.
39. Frith U, Frith CD. Development and neurophysiology of mentalizing. Philos Trans R Soc Lond B Biol Sci. 2003;358:459–73.
40. Abu-Akel A. A neurobiological mapping of theory of mind. Brain Res Rev. 2003;43:29–40.
41. Pike MM, Barnes MA, Barron RW. The role of illustrations in children's inferential comprehension. J Exp Child Psychol. 2010;105:243–55.
42. Steinbeck K, Towns S, Bennett D. Adolescent and young adult medicine is a special and specific area of medical practice. J Paediatr Child Health. 2014; 50:427–31.
43. Dahl RE. Adolescent brain development: a period of vulnerabilities and opportunities. Keynote address. Ann N Y Acad Sci. 2004;1021:1–22.
44. Braams BR, Van Leijenhorst L, Crone EA. Risks, Rewards, and the Developing Brain in Childhood and Adolescence. In: Reyna VF, Zayas V, editors. The neuroscience of risky decision making. Washington DC: American Psychological Association; 2014.
45. Crone EA. Het puberende brein. The Netherlands: Bert Bakker; 2008.
46. Steinberg L. A behavioral scientist looks at the science of adolescent brain development. Brain Cogn. 2010;72:160–4.
47. Casey BJ, Jones RM, Hare TA. The adolescent brain. Ann N Y Acad Sci. 2008; 1124:111–26.
48. Blakemore SJ, Robbins TW. Decision-making in the adolescent brain. Nat Neurosci. 2012;15:1184–91.
49. Van Leijenhorst L, Zanolie K, Van Meel CS, Westenberg PM, Rombouts SA, Crone EA. What motivates the adolescent? Brain regions mediating reward sensitivity across adolescence. Cereb Cortex. 2010;20:61–9.
50. Steinberg L. A social neuroscience perspective on adolescent risk-taking. Dev Rev. 2008;28:78–106.

51. Gogtay N, Giedd JN, Lusk L, Hayashi KM, Greenstein D, Vaituzis AC, Nugent 3rd TF, Herman DH, Clasen LS, Toga AW, et al. Dynamic mapping of human cortical development during childhood through early adulthood. Proc Natl Acad Sci U S A. 2004;101:8174–9.
52. Mills KL, Goddings AL, Clasen LS, Giedd JN, Blakemore SJ. The developmental mismatch in structural brain maturation during adolescence. Dev Neurosci. 2014;36:147–60.
53. Crone EA, Dahl RE. Understanding adolescence as a period of social-affective engagement and goal flexibility. Nat Rev Neurosci. 2012;13:636–50.
54. Blakemore SJ. Imaging brain development: the adolescent brain. Neuroimage. 2012;61:397–406.
55. Blakemore SJ, Mills KL. Is adolescence a sensitive period for sociocultural processing? Annu Rev Psychol. 2014;65:187–207.
56. Albert D, Chein J, Steinberg L. The teenage brain: influences on adolescent decision making. Curr Dir Psychol. 2013;22:114–20.
57. Fischhoff B, Bruine de Bruin W, Parker AM, Millstein SG, Halpern-Felsher BL. Adolescents' perceived risk of dying. J Adolesc Health. 2010;46:265–9.
58. Galvan A. Insights about adolescent behavior, plasticity, and policy from neuroscience research. Neuron. 2014;83:262–5.
59. Martenson EK, Fagerskiold AM. A review of children's decision-making competence in health care. J Clin Nurs. 2008;17:3131–41.
60. De Lourdes LM, Larcher V, Kurz R. Informed consent/assent in children. Statement of the Ethics Working Group of the Confederation of European Specialists in Paediatrics (CESP). Eur J Pediatr. 2003;162:629–33.
61. Kurz R, Gill D, Mjones S. Ethical issues in the daily medical care of children. Eur J Pediatr. 2006;165:83–6.
62. Gill D, Crawley FP, LoGiudice M, Grosek S, Kurz R, de Lourdes-Levy M, Mjones S, Nicolopoulos D, Rubino A, Sauer PJ, et al. Guidelines for informed consent in biomedical research involving paediatric populations as research participants. Eur J Pediatr. 2003;162:455–8.
63. de Vries MC, van Leeuwen E. Ethics of medical scientific research: informed consent and the therapeutic misconception. Ned Tijdschr Geneeskd. 2008;152(12):679–83.
64. Alderson P. Competent children? Minors' consent to health care treatment and research. Soc Sci Med. 2007;65:2272–83.
65. Bos W, Tromp K, Tibboel D, Pinxten W. Ethical aspects of clinical research with minors. Eur J Pediatr. 2013;172:859–66.
66. Nolen-Hoeksema S, Fredrickson BL, Loftus GR, Wagenaar WA. Atkinson & Hilgard's Introduction to Psychology. 15th ed. Hampshire: Wadsworth Cengage Learning; 2009.
67. Friederici AD, Brauer J, Lohmann G. Maturation of the language network: from inter- to intrahemispheric connectivities. PLoS One. 2011;6:e20726.
68. Brauer J, Anwander A, Friederici AD. Neuroanatomical prerequisites for language functions in the maturing brain. Cereb Cortex. 2011;21:459–66.
69. Posner MI, Petersen SE. The attention system of the human brain. Annu Rev Neurosci. 1990;13:25–42.
70. Petersen SE, Posner MI. The attention system of the human brain: 20 years after. Annu Rev Neurosci. 2012;35:73–89.
71. Anderson P. Assessment and development of executive function (EF) during childhood. Child Neuropsychol. 2002;8:71–82.
72. Mezzacappa E. Alerting, orienting, and executive attention: developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. Child Dev. 2004;75:1373–86.
73. Dosenbach NU, Fair DA, Cohen AL, Schlaggar BL, Petersen SE. A dual-networks architecture of top-down control. Trends Cogn Sci. 2008;12:99–105.
74. Rueda MR, Rothbart MK, McCandliss BD, Saccomanno L, Posner MI. Training, maturation, and genetic influences on the development of executive attention. Proc Natl Acad Sci U S A. 2005;102:14931–6.
75. Dionne J, Cadoret G. Development of active controlled retrieval during middle childhood. Dev Psychobiol. 2013;55:443–9.
76. Goldstein G, Allen DN, Thaler NS, Luther JF, Panchalingam K, Pettegrew JW. Developmental aspects and neurobiological correlates of working and associative memory. Neuropsychology. 2014;28:496–505.
77. Parsons LM, Osherson D. New evidence for distinct right and left brain Systems for deductive versus probabilistic reasoning. Cereb Cortex. 2001;11:954–65.
78. Verschueren N, Schaeken W, d'Ydewalle G. Everyday conditional reasoning: a working memory-dependent tradeoff between counterexample and likelihood use. Mem Cognit. 2005;33:107–19.
79. Reyna VF, Brainerd CJ. Dual processes in decision making and developmental neuroscience: a fuzzy-trace model. Dev Rev. 2011;31:180–206.
80. Pillow BH, Hill V, Boyce A, Stein C. Understanding inference as a source of knowledge: children's ability to evaluate the certainty of deduction, perception, and guessing. Dev Psychol. 2000;36:169–79.
81. Pillow BH. Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. Child Dev. 2002;73:779–92.
82. Markovits H, Thompson V. Different developmental patterns of simple deductive and probabilistic inferential reasoning. Mem Cognit. 2008;36:1066–78.
83. Amsterlaw J. Children's beliefs about everyday reasoning. Child Dev. 2006;77:443–64.
84. Cohen MX, Heller AS, Ranganath C. Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. Cogn Brain Res. 2005;23:61–70.
85. Morrongiello BA, Rennie H. Why do boys engage in more risk taking than girls? The role of attributions, beliefs, and risk appraisals. J Pediatr Psychol. 1998;23:33–43.
86. Korkmaz B. Theory of mind and neurodevelopmental disorders of childhood. Pediatr Res. 2011;69:101R–8R.
87. Pfeifer JH, Blakemore SJ. Adolescent social cognitive and affective neuroscience: past, present, and future. Soc Cogn Affect Neurosci. 2012;7:1–10.
88. Schwanenflugel PJ, Henderson RL, Fabricius WV. Developing organization of mental verbs and theory of mind in middle childhood: evidence from extensions. Dev Psychol. 1998;34:512–24.