EXHIBIT 68

Evidence review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria

This document will help inform Dr Hilary Cass' independent review into gender identity services for children and young people. It was commissioned by NHS England and Improvement who commissioned the Cass review. It aims to assess the evidence for the clinical effectiveness, safety and cost-effectiveness of gonadotrophin releasing hormone (GnRH) analogues for children and adolescents aged 18 years or under with gender dysphoria.

The document was prepared by NICE in October 2020.

The content of this evidence review was up to date on 14 October 2020. See summaries of product characteristics (SPCs), British National Formulary (BNF) or the Medicines and Healthcare products Regulatory Agency (MHRA) or NICE websites for up-to-date information.

Contents

1. Introduction

This review aims to assess the evidence for the clinical effectiveness, safety and cost-effectiveness of gonadotrophin releasing hormone (GnRH) analogues for children and adolescents aged 18 years or under with gender dysphoria. The review follows the NHS England Specialised Commissioning process and template and is based on the criteria outlined in the PICO framework (see appendix A). This document will help inform Dr Hilary Cass' independent review into gender identity services for children and young people.

Gender dysphoria in children, also known as gender identity disorder or gender incongruence of childhood (World Health Organisation 2020), refers to discomfort or distress that is caused by a discrepancy between a person's gender identity (how they see themselves[1] regarding their gender) and that person's sex assigned at birth and the associated gender role, and/or primary and secondary sex characteristics (Diagnostic and Statistical Manual of Mental Disorders 2013).

GnRH analogues suppress puberty by delaying the development of secondary sexual characteristics. The intention is to alleviate the distress associated with the development of secondary sex characteristics, thereby providing a time for on-going discussion and exploration of gender identity before deciding whether to take less reversible steps. In England, the GnRH analogue triptorelin (a synthetic decapeptide analogue of natural GnRH, which has marketing authorisations for the treatment of prostate cancer, endometriosis and precocious puberty [onset before 8 years in girls and 10 years in boys]) is used for this purpose. The use of triptorelin for children and adolescents with gender dysphoria is off-label.

For children and adolescents with gender dysphoria it is recommended that management plans are tailored to the needs of the individual, and aim to ameliorate the potentially negative impact of gender dysphoria on general developmental processes, support young people and their families in managing the uncertainties inherent in gender identity development and provide on-going opportunities for exploration of gender identity. The plans may also include psychological support and exploration and, for some individuals, the use of GnRH analogues in adolescence to suppress puberty; this may be followed later with gender-affirming hormones of the desired sex (NHS England 2013).

2. Executive summary of the review

Nine observational studies were included in the evidence review. Five studies were retrospective observational studies (Brik et al. 2020, Joseph et al. 2019, Khatchadourian et al. 2014, Klink et al. 2015, Vlot et al. 2017), 3 studies were prospective longitudinal observational studies (Costa et al. 2015, de Vries et al. 2011, Schagen et al. 2016) and 1 study was a cross-sectional study (Staphorsius et al. 2015). Two studies (Costa et al. 2015

[1] Gender refers to the roles, behaviours, activities, attributes and opportunities that any society considers appropriate for girls and boys, and women and men (World Health Organisation, Health Topics: Gender).

and Staphorsius et al. 2015) provided comparative evidence and the remaining 7 studies used within-person, before and after comparisons.

The terminology used in this topic area is continually evolving and is different depending on stakeholder perspectives. In this evidence review we have used the phrase 'people's assigned sex at birth' rather than natal or biological sex, gonadotrophin releasing hormone (GnRH) analogues rather than 'puberty blockers' and gender-affirming hormones rather than 'cross sex hormones'. The research studies included in this evidence review may use historical terms which are no longer considered appropriate.

In children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Critical outcomes

The critical outcomes for decision making are the impact on gender dysphoria, mental health and quality of life. The quality of evidence for these outcomes was assessed as very low certainty using modified GRADE.

Impact on gender dysphoria

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones does not affect gender dysphoria (measured using the Utrecht Gender Dysphoria Scale [UGDS]). The mean (±SD) gender dysphoria (UGDS) score was not statistically significantly different at baseline compared with follow-up (n=41, 53.20 [±7.91] versus 53.9 [±17.42], p=0.333).

Impact on mental health

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones may reduce depression (measured using the Beck Depression Inventory-II [BDI-II]). The mean [±SD] BDI score was statistically significantly lower (improved) from baseline compared with follow-up (n=41, 8.31 [±7.12] versus 4.95 [±6.72], p=0.004).

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones does not affect anger (measured using the Trait Anger Scale [TPI]). The mean [±SD] anger (TPI) score was not statistically significantly different at baseline compared with follow-up (n=41, 18.29 [±5.54] versus 17.88 [±5.24], p=0.503).

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones does not affect anxiety (measured using the Trait Anxiety Scale [STAI]). The mean [±SD] anxiety (STAI) score was not statistically significantly different at baseline compared with follow-up (n=41, 39.43 [±10.07] versus 37.95 [±9.38], p=0.276).

Impact on quality of life

No evidence was identified.

Important outcomes

The important outcomes for decision making are impact on body image, psychosocial impact, engagement with health care services, impact on extent of and satisfaction with surgery and stopping treatment. The quality of evidence for all these outcomes was assessed as very low certainty using modified GRADE.

Impact on body image

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones does not affect body image (measured using the Body Image Scale [BIS]). The mean [±SD] body image (BIS) scores were not statistically significantly different from baseline compared with follow-up for primary sexual characteristics (n=57, 4.10 [±0.56] versus 3.98 [±0.71], p=0.145), secondary sexual characteristics (n=57, 2.74 [±0.65] versus 2.82 [±0.68], p=0.569) or neutral body characteristics (n=57, 2.41 [±0.63] versus 2.47 [±0.56], p=0.620).

Psychosocial impact

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones may improve psychosocial impact over time (measured using the Children's Global Assessment Scale [CGAS]). The mean [±SD] CGAS score was statistically significantly higher (improved) from baseline compared with follow-up (n=41, 70.24 [±10.12] versus 73.90 [±9.63], p=0.005).

This study also found that psychosocial functioning may improve over time (measured using the Child Behaviour Checklist [CBCL] and the self-administered Youth Self-Report [YSR]). The mean [±SD] CBCL scores were statistically significantly lower (improved) from baseline compared with follow-up for Total T score (n=54, 60.70 [±12.76] versus 54.46 [±11.23], p<0.001), internalising T score (n=54, 61.00 [±12.21] versus 52.17 [±9.81], p<0.001) and externalising T score (n=54, 58.04 [±12.99] versus 53.81 [±11.86], p=0.001). The mean [±SD] YSR scores were statistically significantly lower (improved) from baseline compared with follow-up for Total T score (n=54, 55.46 [±11.56] versus 50.00 [±10.56], p<0.001), internalising T score (n=54, 56.04 [±12.49] versus 49.78 [±11.63], p<0.001) and externalising T score (n=54, 53.30 [±11.87] versus 49.98 [±9.35], p=0.009). The proportion of adolescents scoring in the clinical range decreased from baseline to follow up on the CBCL total problem scale (44.4% versus 22.2%, p=0.001) and the internalising scale of the YSR (29.6% versus 11.1%, p=0.017).

The study by Costa et al. 2015 in 201 adolescents with gender dysphoria who had 6 months of psychological support followed by either GnRH analogues and continued psychological support or continued psychological support only, found that during treatment with GnRH analogues psychosocial impact in terms of global functioning may improve over time (measured using the CGAS). In the group receiving GnRH analogues, the mean [±SD] CGAS score was statistically significantly higher (improved) after 6 months (n=60, 64.70 [±13.34]) and 12 months (n=35, 67.40 [±13.39]) compared with baseline (n=101, 58.72 [±11.38], p=0.003 and p<0.001, respectively). However, there was no statistically significant difference in global functioning (CGAS scores) between the group receiving GnRH analogues plus psychological support and the group receiving psychological support only at any time point.

The study by Staphorsius et al. 2015 in 40 adolescents with gender dysphoria (20 of whom were receiving GnRH analogues) gave mean [±SD] CBCL scores for each group, but statistical analysis is unclear (transfemales receiving GnRH analogues 57.4 [±9.8], transfemales not receiving GnRH analogues 58.2 [±9.3], transmales receiving GnRH analogues 57.5 [±9.4], transmales not receiving GnRH analogues 63.9 [±10.5]).

Engagement with health care services

The study by Brik et al. 2018 in 143 children and adolescents with gender dysphoria receiving GnRH analogues found that 9 adolescents in the original sampling frame (9/214, 4.2%) were excluded from the study because they stopped attending appointments.

The study by Costa et al. 2015 in 201 adolescents with gender dysphoria who had 6 months of psychological support followed by either GnRH analogues and continued psychological support or continued psychological support only had a large loss to follow-up over time. The sample size at baseline and 6 months was 201, which dropped by 39.8% to 121 after 12 months and by 64.7% to 71 at 18 months follow-up. No explanation of the reasons for loss to follow-up are reported.

Impact on extent of and satisfaction with surgery

No evidence was identified.

Stopping treatment

The study by Brik et al. 2018 in 143 children and adolescents with gender dysphoria receiving GnRH analogues reported the reasons for stopping GnRH analogues. During the follow-up period 6.2% (9/143) of adolescents had stopped GnRH analogues after a median duration of 0.8 years (range 0.1 to 3.0). Five adolescents stopped treatment because they no longer wished to receive gender-affirming treatment for various reasons. In 4 adolescents (all transmales), GnRH analogues were stopped mainly because of adverse effects (such as mood and emotional lability), although they wanted to continue treatments for gender dysphoria.

The study by Khatchadourian et al. 2014 in 27 adolescents with gender dysphoria who started GnRH analogues reported the reasons for stopping them. Eleven out of 26 where data was available (42%) stopped GnRH analogues during follow up.

In children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Evidence was available for bone density, cognitive development or functioning, and other safety outcomes. The quality of evidence for all these outcomes was assessed as very low certainty using modified GRADE.

Bone density

The study by Joseph et al. 2019 in 70 adolescents with gender dysphoria found that GnRH analogues may reduce the expected increase in lumbar or femoral bone density (measured with the z-score). However, the z-scores were largely within 1 standard deviation of normal.

and actual lumbar or femoral bone density values were not statistically significantly different between baseline and follow-up:

- The mean z-score [±SD] for lumbar bone mineral apparent density (BMAD) was statistically significantly lower at 1 year compared with baseline in transfemales (baseline 0.859 [±0.154], 1 year −0.228 [±1.027], p=0.000) and transmales (baseline −0.186 [±1.230], 1 year −0.541 [±1.396], p=0.006).
- The mean z-score [±SD] for lumbar BMAD was statistically significantly lower after receiving GnRH analogues for 2 years compared with baseline in transfemales (baseline 0.486 [±0.809], 2 years −0.279 [±0.930], p=0.000) and transmales (baseline −0.361 [±1.439], 2 years −0.913 [±1.318], p=0.001).
- The mean z-score [±SD] for femoral neck bone mineral density (BMD) was statistically significantly lower after receiving GnRH analogues for 2 years compared with baseline in transfemales (baseline 0.0450 [±0.781], 2 years −0.600 [±1.059], p=0.002) and transmales (baseline −1.075 [±1.145], 2 years −1.779 [±0.816], p=0.001).

The study by Klink et al. 2015 in 34 adolescents with gender dysphoria found that GnRH analogues may reduce the expected increase in lumbar (transmales only), but not femoral bone density. However, the z-scores are largely within 1 standard deviation of normal. Actual lumbar or femoral bone density values were not statistically significantly different between baseline and follow-up (apart from BMD measurements in transmales):

- The mean z-score [±SD] for lumbar BMAD was not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales, but was statistically significantly lower when starting gender-affirming hormones in transmales (GnRH analogues 0.28 [±0.90], gender-affirming hormones −0.50 [±0.81], p=0.004).

The study by Vlot et al. 2017 in 70 adolescents with gender dysphoria found that GnRH analogues may reduce the expected increase in lumbar or femoral bone density. However, the z-scores were largely within 1 standard deviation of normal. Actual lumbar or femoral bone density values were not statistically significantly different between baseline and follow-up (apart from in transmales with a bone age ≥14 years). This study reported change in bone density from starting GnRH analogues to starting gender-affirming hormones by bone age:

- The median z-score [range] for lumbar BMAD in transfemales with a bone age of <15 years was statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (GnRH analogues −0.20 [−1.82 to 1.18], gender-affirming hormones −1.52 [−2.36 to 0.42], p=0.001) but was not statistically significantly different in transfemales with a bone age ≥15 years.
- The median z-score [range] for lumbar BMAD in transmales with a bone age of <14 years was statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (GnRH analogues −0.05 [−0.78 to 2.94], gender-affirming hormones −0.84 [−2.20 to 0.87], p=0.003) and in transmales with a bone age ≥14 years (GnRH analogues 0.27 [−1.60 to 1.80], gender-affirming hormones −0.29 [−2.28 to 0.90], p≤0.0001).

- The median z-score [range] for femoral neck BMAD in transfemales with a bone age of <15 years was not statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (GnRH analogues −0.71 [−3.35 to 0.37], gender-affirming hormones −1.32 [−3.39 to 0.21], p≤0.1) or in transfemales with a bone age ≥15 years (GnRH analogues −0.44 [−1.37 to 0.93], gender-affirming hormones −0.36 [−1.50 to 0.46]).
- The z-score for femoral neck BMAD in transmales with a bone age of <14 years was not statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (GnRH analogues −0.01 [−1.30 to 0.91], gender-affirming hormone −0.37 [−2.28 to 0.47]) but was statistically significantly lower in transmales with a bone age ≥14 years (GnRH analogues 0.27 [−1.39 to 1.32], gender-affirming hormones −0.27 [−1.91 to 1.29], p=0.002).

Cognitive development or functioning

The study by Staphorsius et al. 2015 in 40 adolescents with gender dysphoria (20 of whom were receiving GnRH analogues) measured cognitive development or functioning (using an IQ test, and reaction time and accuracy measured using the Tower of London task):

- The mean (±SD) IQ in transfemales receiving GnRH analogues was 94.0 (±10.3) and 109.4 (±21.2) in the control group. In transmales receiving GnRH analogues the mean (±SD) IQ was 95.8 (±15.6) and 98.5 (±15.9) in the control group.
- The mean (±SD) reaction time in transfemales receiving GnRH analogues was 10.9 (±4.1) and 9.9 (±3.1) in the control group. In transmales receiving GnRH analogue it was 9.9 (±3.1) and 10.0 (±2.0) in the control group.
- The mean (±SD) accuracy score in transfemales receiving GnRH analogues was 73.9 (±9.1) and 83.4 (±9.5) in the control group. In transmales receiving GnRH analogues it was 85.7 (±10.5) and 88.8 (±9.7) in the control group.

No statistical analyses or interpretation of the results was reported.

Other safety outcomes

The study by Schagen et al. 2016 in 116 adolescents with gender dysphoria found that GnRH analogues do not affect renal or liver function:

- There was no statistically significant difference between baseline and 1 year results for serum creatinine in transfemales, but there was a statistically significant decrease between baseline and 1 year in transmales (p=0.01).
- Glutamyl transferase, alanine aminotransferase (ALT), and aspartate aminotransferase (AST) levels did not significantly change from baseline to 12 months of treatment.

The study by Khatchadourian et al. 2014 in 27 adolescents with gender dysphoria who started GnRH analogues narratively reported adverse effects from GnRH analogues in 26 adolescents:

- 1 transmale developed sterile abscesses; they were switched from leuprolide acetate to triptorelin, and this was well tolerated
- 1 transmale developed leg pains and headaches, which eventually resolved
- 1 participant gained 19 kg within 9 months of starting GnRH analogues.

In children and adolescents with gender dysphoria, what is the cost-effectiveness of GnRH analogues compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?

No cost-effectiveness evidence was found for GnRH analogues in children and adolescents with gender dysphoria.

From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may benefit from GnRH analogues more than the wider population of interest?

Some studies reported data separately for the following subgroups of children and adolescents with gender dysphoria: sex assigned at birth males (transfemales) and sex assigned at birth females (transmales). This included some direct comparisons of these subgroups, and differences were largely seen at baseline as well as follow up. No evidence was found for other specified subgroups.

Sex assigned at birth males (transfemales)

Impact on gender dysphoria

The study by Costa et al. 2015 in 201 adolescents with gender dysphoria who had 6 months of psychological support followed by either GnRH analogues and continued psychological support or continued psychological support only, found that gender dysphoria (measured using the UGDS) in sex assigned at birth males is lower than in sex assigned at birth females. Sex assigned at birth males had a statistically significantly lower (improved) mean [±SD] UGDS score of 51.6 [±9.7] compared with sex assigned at birth females (56.1 [±4.3], p<0.001), but it was not reported if this was at baseline or follow-up.

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that gender dysphoria (measured using the UGDS) in sex assigned at birth males is lower than in sex assigned at birth females at baseline and follow up. The mean [±SD] UGDS score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean UGDS score: 47.95 [±9.70] versus 56.57 [±3.89]) and follow up (n=not reported, 49.67 [±9.47] versus 56.62 [±4.00]); between sex difference p<0.001).

Impact on mental health

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that the impact on mental health (depression, anger and anxiety) may be different in sex assigned at birth males compared with sex assigned at birth females. Over time there was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for depression, but sex assigned at birth males had statistically significantly lower levels of anger and anxiety than sex assigned at birth females at baseline and follow up.

- The mean [±SD] depression (BDI-II) score was not statistically significantly different in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean BDI score [±SD]: 5.71 [±4.31] versus 10.34 [±8.24]) and follow-up (n=not reported, 3.50 [±4.58] versus 6.09 [±7.93]), between sex difference p=0.057
- The mean [±SD] anger (TPI) score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean TPI score [±SD]: 5.22 [±2.76] versus 6.43 [±2.78]) and follow-

up (n=not reported, 5.00 [±3.07] versus 6.39 [±2.59]), between sex difference p=0.022

- The mean [±SD] anxiety (STAI) score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean STAI score [±SD]: 4.33 [±2.68] versus 7.00 [±2.36]) and follow-up (n=not reported, 4.39 [±2.64] versus 6.17 [±2.69]), between sex difference p<0.001.

Impact on body image

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that the impact on body image may be different in sex assigned at birth males compared with sex assigned at birth females. Sex assigned at birth males are less dissatisfied with their primary and secondary sex characteristics than sex assigned at birth females at both baseline and follow up, but the satisfaction with neutral body characteristics is not different.

- The mean [±SD] BIS score for primary sex characteristics was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean BIS score [±SD]: 4.02 [±0.61] versus 4.16 [±0.52]) and follow up (n=not reported, 3.74 [±0.78] versus 4.17 [±0.58]) between sex difference p=0.047.
- The mean [±SD] BIS score for secondary sex was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean BIS score [±SD]: 2.66 [±0.50] versus 2.81 [±0.76]) and follow up (n=not reported, 2.39 [±0.69] versus 3.18 [±0.42]), between sex difference p=0.001.
- The mean [±SD] BIS score for neutral body characteristics was not statistically significantly different in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, 2.60 [±0.58] versus 2.24 [±0.62], between sex difference p=0.777).

Psychosocial impact

The study by Costa et al. 2015 in 201 adolescents with gender dysphoria who had 6 months of psychological support followed by either GnRH analogues and continued psychological support or continued psychological support only, found that sex assigned at birth males had statistically significant lower mean [±SD] CGAS scores at baseline compared with sex assigned at birth females (n=201, 55.4 [±12.7] versus 59.2 [±11.8], p=0.03), but no conclusions could be drawn.

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that psychosocial impact in terms of global functioning (CGAS) and psychosocial functioning (CBCL and YSR) may be different in sex assigned at birth males compared with sex assigned at birth females, but no conclusions could be drawn.

- There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females (at baseline or follow up) for the CBCL Total T score, the CBCL internalising T score, the YSR Total T score or the YSR internalising T score.

- Sex assigned at birth males had statistically higher mean [±SD] CGAS scores compared with sex assigned at birth females at baseline (n=54, 73.10 [±8.44] versus 67.25 [±11.06]) and follow up (n=54, 77.33 [±8.69] versus 70.30 [±9.44]), between sex difference p=0.021.
- Sex assigned at birth males had statistically lower mean [±SD] CBCL externalising T scores compared with sex assigned at birth females at baseline (n=54, 54.71 [±12.91] versus 60.70 [±12.64]) and follow up (n=54, 48.75 [±10.22] versus 57.87 [±11.66]), between sex difference p=0.015.
- Sex assigned at birth males had statistically lower mean [±SD] YSR externalising T scores compared with sex assigned at birth females at both baseline (n=54, 48.72 [±11.38] versus 57.24 [±10.59]) and follow up (n=54, 46.52 [±9.23] versus 52.97 [±8.51]), between sex difference p=0.004.

Bone density

The studies by Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017 provided evidence on bone density in sex assigned at birth males (see above for details).

Cognitive development or functioning

The study by Staphorsius et al. 2015 provided evidence on cognitive development or functioning in sex assigned at birth males (see above for details).

Other safety outcomes

The study by Schagen et al. 2016 provided evidence on renal function in sex assigned at birth males (see above).

Sex assigned at birth females (transmales)

Impact on gender dysphoria

The studies by de Vries et al. 2011 and Costa et al. 2015 found that gender dysphoria (measured using the UGDS) in sex assigned at birth females is higher than in sex assigned at birth males at baseline and follow up (see above for details).

Impact on mental health

The study by de Vries et al. 2011 found that the impact on mental health (depression, anger and anxiety) may be different in sex assigned at birth females compared with sex assigned at birth males. Over time there was no statistically significant difference between sex assigned at birth females and sex assigned at birth males for depression, but sex assigned at birth females had statistically significantly greater levels of anger and anxiety than sex assigned at birth males at both baseline and follow up (see above for details).

Impact on body image

The study by de Vries et al. 2011 found that the impact on body image may be different in sex assigned at birth females compared with sex assigned at birth males. Sex assigned at birth females are more dissatisfied with their primary and secondary sex characteristics than sex assigned at birth males at both baseline and follow up, but the satisfaction with neutral body characteristics is not different (see above for details).

Psychosocial impact

The studies by de Vries et al. 2011 and Costa et al. 2015 found that psychosocial impact in terms of global functioning (CGAS) and psychosocial functioning (CBCL and YSR) may be different in sex assigned at birth females compared with sex assigned at birth males, but no conclusions could be drawn (see above for details).

Bone density

The studies by Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017 provided evidence on bone density in sex assigned at birth females (see above for details).

Cognitive development or functioning

The study by Staphorsius et al. 2015 provided evidence on cognitive development or functioning in sex assigned at birth females (see above for details).

Other safety outcomes

The study by Schagen et al. 2016 provided evidence on renal function in sex assigned at birth females (see above for details).

From the evidence selected:

(a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?

(b) what were the ages at which participants commenced treatment with GnRH analogues?

(c) what was the duration of treatment with GnRH analogues?

All studies that reported diagnostic criteria for gender dysphoria (6/9 studies) used the version of the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria that was in use at the time. In 5 studies (Costa et al. 2015, Klink et al. 2015, Schagen et al. 2016, Staphorsius et al. 2015 and Vlot et al. 2017) the DSM-fourth edition, text revision (IV-TR) criteria were used. The study by Brik et al. 2020 used DSM-V criteria. It was not reported how gender dysphoria was defined in the remaining 3 studies.

The studies show variation in the age (11 to 18 years old) at which children and adolescents with gender dysphoria started GnRH analogues.

Most studies did not report the duration of treatment with GnRH analogues (Joseph et al. 2019, Khatchadourian et al. 2014, Vlot et al. 2017, Costa et al. 2015, de Vries et al. 2011, Schagen et al. 2016), but where this was reported (Brik et al. 2020, Klink et al. 2015, Staphorsius et al. 2015) there was a wide variation ranging from a few months to about 5 years.

Discussion

A key limitation to identifying the effectiveness and safety of GnRH analogues for children and adolescents with gender dysphoria is the lack of reliable comparative studies. The lack of clear, expected outcomes from treatment with a GnRH analogue (the purpose of which is to suppress secondary sexual characteristics which may cause distress from unwanted pubertal changes) also makes interpreting the evidence difficult.

The studies included in this evidence review are all small, uncontrolled observational studies, which are subject to bias and confounding, and all the results are of very low certainty using modified GRADE. They all reported physical and mental health comorbidities and concomitant treatments very poorly. All the studies are from a limited number of, mainly European, care facilities. They are described as either tertiary referral or expert services but the low number of services providing such care and publishing evidence may bias the results towards the outcomes in these services only and limit extrapolation.

Many of the studies did not report statistical significance or confidence intervals. Changes in outcome scores for clinical effectiveness and bone density were assessed with regards to statistical significance. However, there is relatively little interpretation of whether the changes in outcomes are clinically meaningful.

In the observational, retrospective studies providing evidence on bone density, participants acted as their own controls and change in bone density was determined between starting GnRH analogues and follow up. Observational studies such as these can only show an association with GnRH analogues and bone density; they cannot show that GnRH analogues caused any differences in bone density seen. Because there was no comparator group and participants acted as their own controls, it is not known whether the findings are associated with GnRH analogues or due to changes over time.

Conclusion

The results of the studies that reported impact on the critical outcomes of gender dysphoria and mental health (depression, anger and anxiety), and the important outcomes of body image and psychosocial impact (global and psychosocial functioning), in children and adolescents with gender dysphoria are of very low certainty using modified GRADE. They suggest little change with GnRH analogues from baseline to follow-up.

Studies that found differences in outcomes could represent changes that are either of questionable clinical value, or the studies themselves are not reliable and changes could be due to confounding, bias or chance. It is plausible, however, that a lack of difference in scores from baseline to follow-up is the effect of GnRH analogues in children and adolescents with gender dysphoria, in whom the development of secondary sexual characteristics might be expected to be associated with an increased impact on gender dysphoria, depression, anxiety, anger and distress over time without treatment. The study by de Vries et al. 2011 reported statistically significant reductions in the Child Behaviour Checklist (CBCL) and Youth Self-Report (YSR) scores from baseline to follow up, which include measures of distress. As the aim of GnRH analogues is to reduce distress caused by the development of secondary sexual characteristics, this may be an important finding. However, as the studies all lack appropriate controls who were not receiving GnRH analogues, any positive changes could be a regression to mean.

The results of the studies that reported bone density outcomes suggest that GnRH analogues may reduce the expected increase in bone density (which is expected during puberty). However, as the studies themselves are not reliable, the results could be due to confounding, bias or chance. While controlled trials may not be possible, comparative studies are needed to understand this association and whether the effects of GnRH analogues on bone density are seen after they are stopped. All the studies that reported safety outcomes provided very low certainty evidence.

No cost-effectiveness evidence was found to determine whether or not GnRH analogues are cost-effective for children and adolescents with gender dysphoria.

The results of the studies that reported outcomes for subgroups of children and adolescents with gender dysphoria, suggest there may be differences between sex assigned at birth males (transfemales) and sex assigned at birth females (transmales).

3. Methodology

Review questions

The review question(s) for this evidence review are:

1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
2. For children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
3. For children and adolescents with gender dysphoria, what is the cost-effectiveness of GnRH analogues compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?
4. From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may derive more (or less) advantage from treatment with GnRH analogues than the wider population of children and adolescents with gender dysphoria?
5. From the evidence selected,
   a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?
   b) what were the ages at which participants commenced treatment with GnRH analogues?
   c) what was the duration of treatment with GnRH analogues?

See appendix A for the full review protocol.

Review process

The methodology to undertake this review is specified by NHS England in their 'Guidance on conducting evidence reviews for Specialised Services Commissioning Products' (2020).

The searches for evidence were informed by the PICO document and were conducted on 23 July 2020

See appendix B for details of the search strategy.

Results from the literature searches were screened using their titles and abstracts for relevance against the criteria in the PICO framework. Full text references of potentially

relevant evidence were obtained and reviewed to determine whether they met the inclusion criteria for this evidence review.

See appendix C for evidence selection details and appendix D for the list of studies excluded from the review and the reasons for their exclusion.

Relevant details and outcomes were extracted from the included studies and were critically appraised using a checklist appropriate to the study design. See appendices E and F for individual study and checklist details.

The available evidence was assessed by outcome for certainty using modified GRADE. See appendix G for GRADE Profiles.

4. Summary of included studies

Nine observational studies were identified for inclusion. Five studies were retrospective observational studies (Brik et al. 2020, Joseph et al. 2019, Khatchadourian et al. 2014, Klink et al. 2015, Vlot et al. 2017), 3 studies were prospective longitudinal observational studies (Costa et al. 2015, de Vries et al. 2011, Schagen et al. 2016) and 1 study was a cross-sectional study (Staphorsius et al. 2015).

The terminology used in this topic area is continually evolving and is different depending on stakeholder perspectives. In this evidence review we have used the phrase 'people's assigned sex at birth' rather than natal or biological sex, gonadotrophin releasing hormone (GnRH) analogues rather than 'puberty blockers' and gender-affirming hormones rather than 'cross sex hormones'. The research studies included in this evidence review may use historical terms which are no longer considered appropriate.

Table 1 provides a summary of these included studies and full details are given in appendix E.

Table 1 Summary of included studies

Study	Population	Intervention and comparison	Outcomes reported
Brik et al. 2020 Retrospective observational single-centre study Netherlands	The study was conducted at the Curium-Leiden University Medical Centre gender clinic in Leiden, the Netherlands and involved adolescents with gender dysphoria. The sample size was 143 adolescents (median age at start of treatment was 15.0 years, range 11.1 to 18.6 years in transfemales; 16.1 years, range 10.1 to 17.9 years in transmales) from a sampling frame of 269 children and adolescents registered at the clinic between November 2010 and January 2018.	**Intervention** 143 children and adolescents receiving GnRH analogues (no specific treatment, dose, route or frequency of administration reported). The median duration was 2.1 years (range 1.6–2.8 years). **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Stopping treatment

Study	Population	Intervention and comparison	Outcomes reported
	Participants were included in the study if they were diagnosed with gender dysphoria according to the DSM-5 criteria, registered at the clinic, were prepubertal and within the appropriate age range, and had started GnRH analogues. No concomitant treatments were reported.		
Costa et al. 2015 Prospective longitudinal observational single centre cohort study United Kingdom	The study was conducted at the Gender Identity Development Service in London and involved adolescents with gender dysphoria. The sample size was 201 adolescents (mean [±SD] age 15.52±1.41 years, range 12 to 17 years) from a sampling frame of 436 consecutive adolescents referred to the service between 2010 and 2014. The mean [±SD] age at the start of GnRH analogues was 16.48 [±1.26] years, range 13 to 17 years. Participants were invited to participate following a 6-month diagnostic process using DSM-IV-TR criteria. No concomitant treatments were reported.	**Intervention** 101 adolescents assessed as being immediately eligible for GnRH analogues (no specific treatment, dose or route of administration reported) plus psychological support. The average duration of treatment was approximately 12 months (no exact figure given). **Comparison** 100 adolescents assessed as not immediately eligible for GnRH analogues (more time needed to make the decision to start GnRH analogues) who had psychological support only. None received GnRH analogues throughout the study.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Psychosocial impact
de Vries et al. 2011 Prospective longitudinal observational single centre before and after study Netherlands	The study was conducted at the Amsterdam gender identity clinic of the VU University Medical Centre and involved adolescents who were defined as "transsexual". The sample size was 70 adolescents receiving GnRH analogues (mean age [±SD] at assessment 13.6±1.8 years) from a sampling frame of 196 consecutive adolescents referred to the service between 2000 and 2008. Participants were invited to participate if they subsequently started gender-affirming hormones between 2003 and 2009. No diagnostic criteria or concomitant treatments were reported.	**Intervention** 70 individuals assessed at baseline (T0) before the start of GnRH analogues (no specific treatment, dose or route of administration reported). **Comparison** No comparator.	**Critical Outcomes** • Gender dysphoria • Mental health (depression, anger and anxiety) **Important outcomes** • Body image • Psychosocial impact

Study	Population	Intervention and comparison	Outcomes reported
Joseph et al. 2019 Retrospective longitudinal observational single centre study United Kingdom	This study was conducted at the Early intervention clinic at University College London Hospital (all participants had been seen at the Gender Identity Development Service in London) and involved adolescents with gender dysphoria. The sample size was 70 adolescents with gender dysphoria (no diagnostic criteria described) all offered GnRH analogues. The mean age at the start of treatment was 13.2 years (SD ±1.4) for transfemales and 12.6 years (SD ±1.0) for transmales. Details of the sampling frame were not reported. Further details of how the sample was drawn are not reported. No concomitant treatments were reported.	**Intervention** GnRH analogues. No specific treatment, duration, dose or route of administration reported. **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Safety: bone density
Khatchadourian et al. 2014 Retrospective observational chart review single centre study Canada	This study was conducted at the Endocrinology and Diabetes Unit at British Columbia Children's Hospital, Canada and involved youths with gender dysphoria. The sample size was 27 young people with gender dysphoria who started GnRH analogues (at mean age 14.7 [SD ±1.9] years) out of 84 young people seen at the unit between 1998 and 2011. Diagnostic criteria and concomitant treatments were not reported.	**Intervention** 84 young people with gender dysphoria. For GnRH analogues no specific treatment, duration, dose or route of administration reported. **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Stopping treatment • Safety: adverse effects
Klink et al. 2015 Retrospective longitudinal observational single centre study Netherlands	This study was conducted in the Netherlands at a tertiary referral centre. It is unclear which centre this was. The sample size was 34 adolescents (mean age 14.9 [SD ±1.9] years for transfemales and 15.0 [SD ±2.0] years for transmales at start of GnRH analogues). Details of the sampling frame are not reported. Participants were included if they met DSM-IV-TR criteria for gender identity disorder of adolescence and had been treated with GnRH analogues and gender-affirming hormones during their pubertal years. No concomitant treatments were reported.	**Intervention** The intervention was GnRH analogue monotherapy (triptorelin 3.75 mg subcutaneously every 4 weeks) followed by gender-affirming hormones with discontinuation of GnRH analogues after gonadectomy. Duration of GnRH analogues was 1.3 years (range 0.5 to 3.8 years) in transfemales and 1.5 years (0.25 to 5.2 years in transmales. **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Safety: bone density

Study	Population	Intervention and comparison	Outcomes reported
Schagen et al. 2016 Prospective longitudinal study Netherlands	This study was conducted at the Centre of Expertise on Gender Dysphoria at the VU University Medical Centre (Amsterdam, Netherlands) and involved adolescents with gender dysphoria. The sample size was 116 adolescents (median age [range] 13.6 years [11.6 to 17.9] in transfemales and 14.2 years [11.1 to 18.6] in transmales during first year of GnRH analogues) out of 128 adolescents who started GnRH analogues. Participants were included if they met DSM-IV-TR criteria for gender dysphoria, had lifelong extreme gender dysphoria, were psychologically stable and were living in a supportive environment. No concomitant treatments were reported.	**Intervention** The intervention was GnRH analogue monotherapy (triptorelin 3.75 mg at 0, 2 and 4 weeks followed by intramuscular injections every 4 weeks, for at least 3 months). **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Safety: liver and renal function.
Staphorsius et al. 2015 Cross-sectional (single time point) assessment single centre study Netherlands	This study was conducted at the VU University Medical Centre (Amsterdam, Netherlands) and involved adolescents with gender dysphoria. The sample size was 85, of whom 40 were adolescents with gender dysphoria (20 of whom were being treated with GnRH analogues) and 45 were controls without gender dysphoria (not further reported here). Mean (±SD) age 15.1 (±2.4) years in transfemales and 15.8 (±1.9) years in transmales. Details of the sampling frame are not reported. Participants were included if they were diagnosed with Gender Identity Disorder according to the DSM-IV-TR and at least 12 years old and Tanner stage of at least B2 or G2 to G3 with measurable oestradiol and testosterone levels in girls and boys, respectively. No concomitant treatments were reported.	**Intervention** The intervention was a GnRH analogue (triptorelin 3.75 mg every 4 weeks subcutaneously or intramuscularly). The mean duration of treatment was 1.6 years (SD ±1.0). **Comparison** Adolescents with gender dysphoria not treated with GnRH analogues.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Psychosocial impact • Safety, cognitive functioning
Vlot et al. 2017 Retrospective observational data analysis study Netherlands	This study was conducted at the VU University Medical Centre (Amsterdam, Netherlands) and involved adolescents with gender dysphoria. The sample size was 70 adolescents (median age [range] 15.1 years [11.7 to 18.6] for transmales and 13.5 years [11.5 to	**Intervention** The intervention was a GnRH analogue (triptorelin 3.75 mg every 4 weeks subcutaneously). **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes**

Study	Population	Intervention and comparison	Outcomes reported
	18.3] for transfemales at start of GnRH analogues). Details of the sampling frame are not reported. Participants were included if they had a diagnosis of gender dysphoria according to DSM-IV-TR criteria who were receiving GnRH analogues and then gender-affirming hormones. No concomitant treatments were reported.		• Safety: bone density

Abbreviations: DSM-IV-TR, Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision; GnRH, Gonadotrophin releasing hormone; SD, Standard deviation.

5. Results

In children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement
Clinical Effectiveness	
Critical outcomes	
Impact on gender dysphoria **Certainty of evidence: very low**	This is a critical outcome because gender dysphoria in children and adolescents is associated with significant distress and problems with functioning. One uncontrolled, prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on gender dysphoria in adolescents, measured using the Utrecht Gender Dysphoria Scale (UGDS). The UGDS is a validated screening tool for both adolescents and adults to assess gender dysphoria. It consists of 12 items, to be answered on a 1- to 5-point scale, resulting in a sum score between 12 and 60. The higher the UGDS score the greater the gender dysphoria. The study measured the impact on gender dysphoria at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) UGDS score was not statistically significantly different at baseline compared with follow-up (n=41, 53.20 [±7.91] versus 53.9 [±17.42], p=0.333) **(VERY LOW)**. **This study provides very low certainty evidence that treatment with GnRH analogues, before starting gender-affirming hormones, does not affect gender dysphoria.**

Impact on mental health: depression **Certainty of evidence: very low**	This is a critical outcome because self-harm and thoughts of suicide have the potential to result in significant physical harm and, for completed suicides, the death of the young person. One uncontrolled, prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on depression in children and adolescents with gender dysphoria. Depression was measured using the Beck Depression Inventory-II (BDI-II). The BDI-II is a valid, reliable, and widely used tool for assessing depressive symptoms. There are no specific scores to categorise depression severity, but it is suggested that 0 to 13 is minimal symptoms, 14 to 19 is mild depression, 20 to 28 is moderate depression, and severe depression is 29 to 63. The study provided evidence for depression measured at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) depression (BDI) score was statistically significantly lower (improved) from baseline compared with follow-up (n=41, 8.31 [±7 12] versus 4.95 [±6.72], p=0.004) **(VERY LOW)**. **This study provides very low certainty evidence that treatment with GnRH analogues, before starting gender-affirming hormones, may reduce depression.**
Impact on mental health: anger **Certainty of evidence: very low**	This is a critical outcome because self-harm and thoughts of suicide have the potential to result in significant physical harm and, for completed suicides, the death of the young person. One uncontrolled, prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on anger in children and adolescents with gender dysphoria. Anger was measured using the Trait Anger Scale of the State-Trait Personality Inventory (TPI). This is a validated 20-item inventory tool which measures the intensity of anger as the disposition to experience angry feelings as a personality trait. Higher scores indicate greater anger. The study provided evidence for anger measured at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) anger (TPI) score was not statistically significantly different at baseline compared with follow-up (n=41, 18.29 [±5.54] versus 17.88 [±5.24], p=0.503) **(VERY LOW)**. **This study provides very low certainty evidence that treatment with GnRH analogues, before starting gender-affirming hormones, does not affect anger.**
Impact on mental health: anxiety	This is a critical outcome because self-harm and thoughts of suicide have the potential to result in significant physical harm and, for completed suicides, the death of the young person.

Certainty of evidence: very low	One uncontrolled, prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on anxiety in children and adolescents with gender dysphoria. Anxiety was measured using the Trait Anxiety Scale of the State-Trait Personality Inventory (STAI). This is a validated and commonly used measure of trait and state anxiety. It has 20 items and can be used in clinical settings to diagnose anxiety and to distinguish it from depressive illness. Higher scores indicate greater anxiety. The study provided evidence for anxiety at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) anxiety (STAI) score was not statistically significantly different at baseline compared with follow-up (n=41, 39.43 [±10.07] versus 37.95 [±9.38], p=0.276) (**VERY LOW**). **This study provides very low certainty evidence that treatment with GnRH analogues, before starting gender-affirming hormones, does not affect levels of anxiety.**
Quality of life	This is a critical outcome because gender dysphoria in children and adolescents may be associated with a significant reduction in health-related quality of life. No evidence was identified.
Important outcomes	
Impact on body image **Certainty of evidence: very low**	This is an important outcome because some children and adolescents with gender dysphoria may want to take steps to suppress features of their physical appearance associated with their sex assigned at birth or accentuate physical features of their desired gender. One uncontrolled, prospective observational longitudinal study provided evidence relating to the impact on body image (de Vries et al. 2011). Body image was measured using the Body Image Scale (BIS) which is a validated 30-item scale covering 3 aspects: primary, secondary and neutral body characteristics. Higher scores represent a higher degree of body dissatisfaction. The study (de Vries et al. 2011) provided evidence for body image measured at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) body image (BIS) scores for were not statistically significantly different from baseline compared with follow-up for: • primary sexual characteristics (n=57, 4.10 [±0.56] versus 3.98 [±0.71], p=0.145) • secondary sexual characteristics (n=57, 2.74 [±0.65] versus 2.82 [±0.68], p=0.569)

	• neutral body characteristics (n=57, 2.41 [±0.63] versus 2.47 [±0.56], p=0.620) **(VERY LOW)**. **This study provides very low certainty evidence that treatment with GnRH analogues, before starting gender affirming hormones, does not affect body image.**
Psychosocial impact: global functioning **Certainty of evidence: very low**	This is an important outcome because gender dysphoria in children and adolescents is associated with internalising and externalising behaviours, and emotional and behavioural problems which may impact on social and occupational functioning. One uncontrolled, observational, prospective cohort study (de Vries et al 2011) and one prospective cross-sectional cohort study (Costa et al. 2015) provided evidence relating to psychosocial impact in terms of global functioning. Global functioning was measured using the Children's Global Assessment Scale (CGAS). The CGAS tool is a validated measure of global functioning on a single rating scale from 1 to 100. Lower scores indicate poorer functioning. One study (de Vries et al. 2011) provided evidence for global functioning (CGAS) at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) CGAS score was statistically significantly higher (improved) from baseline compared with follow-up (n=41, 70.24 [±10.12] versus 73.90 [±9.63], p=0.005) **(VERY LOW)**. One study (Costa et al. 2015) in adolescents with gender dysphoria who had 6 months of psychological support followed by either GnRH analogues and continued psychological support (the immediately eligible group) or continued psychological support only (the delayed eligible group who did not receive GnRH analogues) provided evidence for global functioning (CGAS) measured at 4 time points: • at baseline (T0) in both groups, • after 6 months of psychological support in both groups (T1), • after 6 months of GnRH analogues and 12 months of psychological support in the immediately eligible group and 12 months of psychological support only in the delayed eligible group (T2), and • after 18 months of psychological support and 12 months of GnRH analogues in the immediately eligible group and after 18 months of psychological support only in the delayed eligible group (T3). The mean [±SD] CGAS score was statistically significantly higher (improved) for all adolescents (including those not receiving GnRH analogues) at T1, T2 or T3 compared with baseline (T0). For the immediately eligible group (who received GnRH analogues) versus the delayed eligible group (who did not receive GnRH analogues) there were no statistically significant differences in CGAS

	scores between the 2 groups at baseline T0 (n=201, p=0.23), T1 (n=201, p=0.73), T2 (n=121, p=0.49) or T3 (n=71, p=0.14) time points. For the immediately eligible group (who received GnRH analogues), the mean (±SD) CGAS score was not statistically significantly different at: • T1 compared with T0 • T2 compared with T1 • T3 compared with T2. The mean (±SD) CGAS score was statistically significantly higher (improved) at: • T2 compared with T0 (n=60, 64.70 [±13.34] versus n=101, 58.72 [±11.38], p=0.003) • T3 compared with T0 (n=35, 67.40 [±13.39] versus n=101, 58.72 [±11.38], p<0.001) • T3 compared with T1 (n=35, 67.40 [±13.93] versus n=101, 60.89 [±12.17], p<0.001) **(VERY LOW)**. **These studies provide very low certainty evidence that during treatment with GnRH analogues, global functioning may improve over time. However, there was no statistically significant difference in global functioning between GnRH analogues plus psychological support compared with psychological support only at any time point.**
Psychosocial impact: psychosocial functioning **Certainty of evidence: very low**	This is an important outcome because gender dysphoria in children and adolescents is associated with internalising and externalising behaviours, and emotional and behavioural problems which may impact on social and occupational functioning. Two studies provided evidence for this outcome. One uncontrolled, observational, prospective cohort study (de Vries et al. 2011) and 1 cross-sectional observational study (Staphorsius et al. 2015) assessed psychosocial functioning using the Child Behaviour Checklist (CBCL) and the self-administered Youth Self-Report (YSR). The CBCL is a checklist parents complete to detect emotional and behavioural problems in children and adolescents. YSR is similar but is self-completed by the child or adolescent. The scales consist of a Total problems score, which is the sum of the scores of all the problem items. An internalising problem scale sums the anxious/depressed, withdrawn-depressed, and somatic complaints scores while the externalising problem scale combines rule-breaking and aggressive behaviour. The standard scores are scaled so that 50 is average for the child or adolescent's age and gender, with a SD of 10 points. Higher scores indicate greater problems, with a T-score above 63 considered to be in the clinical range. One study (de Vries et al. 2011) provided evidence for psychosocial functioning (CBCL and YSR scores) at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years).

	At follow up, the mean (±SD) CBCL scores were statistically significantly lower (improved) compared with baseline for: • Total T score (n=54, 60.70 [±12.76] versus 54.46 [±11.23], p<0.001 • Internalising T score (n=54, 61.00 [±12.21] versus 52.17 [±9.81], p<0.001) • Externalising T score (n=54, 58.04 [±12.99] versus 53.81 [±11.86], p=0.001). At follow up, the mean (±SD) YSR scores were statistically significantly lower (improved) compared with baseline for: • Total T score (n=54, 55.46 [±11.56] versus 50.00 [±10.56], p<0.001) • Internalising T score (n=54, 56.04 [±12.49] versus 49.78 [±11.63], p<0.001) • Externalising T score (n=54, 53.30 [±11.87] versus 49.98 [±9.35], p=0.009). The proportion of adolescents scoring in the clinical range decreased from baseline to follow up on the CBCL total problem scale (44.4% versus 22.2%, p=0.001) and the internalising scale of the YSR (29.6% versus 11.1%, p=0.017) **(VERY LOW)**. One study (Staphorsius et al. 2015) assessed CBCL in a cohort of adolescents with gender dysphoria (transfemale: n=18, mean [±SD] age 15.1 [±2.4] years and transmale: n=22, mean [±SD] age 15.8 [±1.9] years) either receiving GnRH analogues (transfemale, n=8 and transmale, n=12), or not receiving GnRH analogues (transfemale, n=10 and transmale, n=10). The mean (±SD) CBCL scores for each group were (statistical analysis unclear): • transfemales (total) 57.8 [±9.2] • transfemales receiving GnRH analogues 57.4 [±9.8] • transfemales not receiving GnRH analogues 58.2 [±9.3] • transmales (total) 60.4 [±10.2] • transmales receiving GnRH analogues 57.5 [±9.4] • transmales not receiving GnRH analogues 63.9 [±10.5] **(VERY LOW)**. **These studies provide very low certainty evidence that during treatment with GnRH analogues psychosocial functioning may improve, with the proportion of adolescents in the clinical range for some CBCL and YSR scores decreasing over time.**
Engagement with health care services **Certainty of evidence: very low**	This is an important outcome because patient engagement with health care services will impact on their clinical outcomes. Two uncontrolled observational cohort studies provided evidence relating to loss to follow up, which could be a marker of engagement with health care services (Brik et al. 2018 and Costa et al. 2015). In one retrospective study (Brik et al. 2018), 9 adolescents (9/214, 4.2%) who had stopped attending appointments were excluded from the study between November 2010 and July 2019 **(VERY LOW)**.

	One prospective study (Costa et al. 2015) had evidence for a large loss to follow-up over time. The sample size at baseline (T0) and 6 months (T1) was 201, which dropped by 39.8% to 121 after 12 months (T2) and by 64.7% to 71 at 18 months follow-up (T3). No explanation of the reasons for loss to follow-up are reported **(VERY LOW)**. Due to their design there was no reported loss to follow-up in the other 3 effectiveness studies (de Vries et al 2011; Khatchadourian et al. 2014; Staphorsius et al. 2015). **These studies provide very low certainty evidence about loss to follow up, which could be a marker of engagement with health care services, during treatment with GnRH analogues. Due to the large variation in rates between studies no conclusions could be drawn.**
Impact on extent of and satisfaction with surgery	This is an important outcome because some children and adolescents with gender dysphoria may proceed to transitioning surgery. No evidence was identified.
Stopping treatment **Certainty of evidence: very low**	This is an important outcome because there is uncertainty about the short- and long-term safety and adverse effects of GnRH analogues in children and adolescents with gender dysphoria. Two uncontrolled, retrospective, observational cohort studies provided evidence relating to stopping GnRH analogues. One study had complete reporting of the cohort (Brik et al. 2018), the other (Khatchadourian et al. 2014) had incomplete reporting of its cohort, particularly for transfemales where outcomes for only 4/11 were reported. Brik et al. 2018 narratively reported the reasons for stopping GnRH analogues in a cohort of 143 adolescents (38 transfemales and 105 transmales). Median age at the start of GnRH analogues was 15.0 years (range, 11.1–18.6 years) in transfemales and 16.1 years (range, 10.1–17.9 years) in transmales. Of these adolescents, 125 (87%, 36 transfemales, 89 transmales) subsequently started gender-affirming hormones after 1.0 (0.5–3.8) and 0.8 (0.3–3.7) years of GnRH analogues. At the time of data collection, the median duration of GnRH analogue use was 2.1 years (1.6–2.8). During the follow-up period 6.3% (9/143) of adolescents had discontinued GnRH analogues after a median duration of 0.8 years (range 0.1 to 3.0). The percentages and reasons for stopping were: • 2.8% (4/143) stopped GnRH analogues although they wanted to continue endocrine treatments for gender dysphoria: ○ 1 transmale stopped due to increase in mood problems, suicidal thoughts and confusion attributed to GnRH analogues ○ 1 transmale had hot flushes, increased migraines, fear of injections, stress at school and unrelated medical issues, and temporarily stopped treatment (after 4 months) and restarted 5 months later. ○ 1 transmale had mood swings 4 months after starting GnRH analogues. After 2.2 years had unexplained severe nausea and rapid weight loss and discontinued GnRH analogues after 2.4 years

	○ 1 transmale stopped GnRH analogues because of inability to regularly collect medication and attend appointments for injections. • 3.5% (5/143) stopped treatment because they no longer wished to receive gender-affirming treatment for various reasons **(VERY LOW)**. Khatchadourian et al. 2014 narratively reported the reasons for stopping GnRH analogues in a cohort of 26 adolescents (15 transmales and 11 transfemales). 42% (11/26) discontinued GnRH analogues during follow-up between 1998 and 2011. Of 15 transmales receiving GnRH analogues, 14 received testosterone during the observation period, of which: • 7 continued GnRH analogues after starting testosterone • 7 stopped GnRH analogues after a median of 3.0 years (range 0.2 to 9.2 years), of which: ○ 5 stopped after hysterectomy and salpingo-oophorectomy ○ 1 stopped after 2.2 years (transitioned to gender-affirming hormones) ○ 1 stopped after <2 months due to mood and emotional lability **(VERY LOW)**. Of 11 transfemales receiving GnRH analogues, 5 received oestrogen during the observation period, of which: • 4 continued GnRH analogues after starting oestrogen • 1 stopped GnRH analogues when taking oestrogen (no reason reported) **(VERY LOW)**. Of the remaining 6 transfemales taking GnRH analogues: • 1 stopped GnRH analogues after a few months due to emotional lability • 1 stopped GnRH analogues before taking oestrogen (the following year delayed due to heavy smoking) • 1 stopped GnRH analogues after 13 months due not to pursuing transition **(VERY LOW)**. **These studies provide very low certainty evidence for the number of adolescents who stop GnRH analogues and the reasons for this.**

Abbreviations: GnRH, gonadotrophin releasing hormone; SD, standard deviation.

In children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement
Safety	
Change in bone density: lumbar	This is an important outcome because puberty is an important time for bone development and puberty suppression may affect bone development, as shown by changes in lumbar bone density.

Certainty of evidence: very low	Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on bone density (based on lumbar BMAD) between starting with a GnRH analogue and at 1 and 2 year intervals (Joseph et al. 2019), and between starting GnRH analogues and starting gender-affirming hormones (Klink et al. 2015 and Vlot et al. 2017). All outcomes were reported separately for transfemales and transmales; also see subgroups table below. BMAD is a size adjusted value of BMD incorporating body size measurements using UK norms in growing adolescents. It was reported as g/cm^3 and as z-scores. Z-scores report how many standard deviations from the mean a measurement sits. A z-score of 0 is equal to the mean, a z-score of −1 is equal to 1 standard deviation below the mean, and a z-score of +1 is equal to 1 standard deviation above the mean. One retrospective observational study (Joseph et al. 2019, n=70) provided non-comparative evidence on change in lumbar BMAD increase using z-scores. • The z-score for lumbar BMAD was statistically significantly lower at 2 years compared with baseline in transfemales (z-score [±SD]: baseline 0.486 [0.809], 2 years −0.279 [0.930], p=0.000) and transmales (baseline −0.361 [1.439], 2 years −0.913 [1.318], p=0.001) **(VERY LOW)**. • The z-score for lumbar BMAD was statistically significantly lower at 1 year compared with baseline in transfemales (baseline 0.859 [0.154], 1 year −0.228 [1.027], p=0.000) and transmales (baseline −0.186 [1.230], 1 year −0.541 [1.396], p=0.006) **(VERY LOW)**. • Actual lumbar BMAD values in g/cm^3 were not statistically significantly different between baseline and 1 or 2 years in transfemales or transmales **(VERY LOW)**. Two retrospective observational studies (Klink et al. 2015 and Vlot et al. 2017, n=104 in total) provided non-comparative evidence on change in lumbar BMAD between starting GnRH analogues and starting gender-affirming hormones. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. In Klink et al. 2015 the z-score for lumbar BMAD was not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales but was statistically significantly lower when starting gender-affirming hormones in transmales (z-score mean [±SD]: GnRH analogue 0.28 [±0.90], gender-affirming hormone −0.50 [±0.81], p=0.004). Actual lumbar BMAD values in g/cm^3 were not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales or transmales **(VERY LOW)**. Vlot et al. 2017 reported change from starting GnRH analogues to starting gender-affirming hormones in lumbar BMAD by bone age. • The z-score for lumbar BMAD in transfemales with a bone age of <15 years was statistically significantly lower at starting gender-affirming hormone treatment than at starting GnRH

	analogues (z-score median [range]: GnRH analogue −0.20 [−1.82 to 1.18], gender-affirming hormone −1.52 [−2.36 to 0.42], p=0.001) but was not statistically significantly different in transfemales with a bone age ≥15 years **(VERY LOW)**. • The z-score for lumbar BMAD in transmales with a bone age of <14 years was statistically significantly lower at starting gender-affirming hormone treatment than at starting GnRH analogues (z-score median [range]: GnRH analogue −0.05 [−0.78 to 2.94], gender-affirming hormone −0.84 [−2.20 to 0.87], p=0.003) and in transmales with a bone age ≥14 years (GnRH analogue 0.27 [−1.60 to 1.80], gender-affirming hormone −0.29 [−2.28 to 0.90], p≤0.0001) **(VERY LOW)**. • Actual lumbar BMAD values in g/cm^3 were not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales or transmales with young or old bone age **(VERY LOW)**. Two uncontrolled, observational, retrospective studies provided evidence for the effect of GnRH analogues on bone density (based on lumbar BMD) between starting GnRH analogues and either at 1 or 2 year intervals (Joseph et al. 2019), or starting gender-affirming hormones (Klink et al. 2015). All outcomes were reported separately for transfemales and transmales, also see subgroups table below. One retrospective observational study (Joseph et al. 2019, n=70) provided non-comparative evidence on change in lumbar BMD increase using z-scores. • The z-score for lumbar BMD was statistically significantly lower at 2 years compared with baseline in transfemales (z-score mean [±SD]: baseline 0.130 [0.972], 2 years −0.890 [±1.075], p=0.000) and transmales (baseline −0.715 [±1.406], 2 years −2.000 [1.384], p=0.000) **(VERY LOW)**. • The z-score for lumbar BMD was statistically significantly lower at 1 year compared with baseline in transfemales (z-score mean [±SD]: baseline −0.016 [±1.106], 1 year −0.461 [±1.121], p=0.003) and transmales (baseline −0.395 [±1.428], 1 year −1.276 [±1.410], p=0.000) **(VERY LOW)**. • With the exception of transmales, where lumbar BMD in kg/m^2 increased between baseline and 1 year (mean [±SD]: baseline 0.694 [±0.149], 1 year 0.718 [±0.124], p=0.006), actual lumbar BMD values were not statistically significantly different between baseline and 1 or 2 years in transfemales or between 0 and 2 years in transmales **(VERY LOW)**. One retrospective observational study (Klink et al. 2015, n=34) provided non-comparative evidence on change in lumbar BMD between starting GnRH analogues and starting gender-affirming hormones. • The z-score for lumbar BMD was not statistically significantly different between starting GnRH analogue and starting gender-affirming hormone treatment in transfemales, but was statistically significantly lower when starting gender-affirming hormones in transmales (z-score mean [±SD]: GnRH analogue 0.17 [±1.18], gender-affirming hormone −0.72 [±0.99], p<0.001) **(VERY LOW)**.

	• Actual lumbar BMD in g/cm² was not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales but was statistically significantly lower when starting gender-affirming hormones in transmales (mean [±SD]: GnRH analogues 0.95 [±0.12], gender-affirming hormones 0.91 [±0.10], p=0.006) **(VERY LOW)**. **These studies provide very low certainty evidence that GnRH analogues reduce the expected increase in lumbar bone density (BMAD or BMD) compared with baseline (although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically significantly decrease actual lumbar bone density (BMAD or BMD).**
Change in bone density: femoral **Certainty of evidence: very low**	This is an important outcome because puberty is an important time for bone development and puberty suppression may affect bone development, as shown by changes in femoral bone density. Two uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on bone density (based on femoral BMAD) between starting treatment with a GnRH analogue and starting gender-affirming hormones (Klink et al. 2015 and Vlot et al. 2017). All outcomes were reported separately for transfemales and transmales; also see subgroups table below. One retrospective observational study (Klink et al. 2015, n=34) provided non-comparative evidence on change in femoral area BMAD between starting GnRH analogues and starting gender-affirming hormones. All outcomes were reported separately for transfemales and transmales. • The z-score for femoral area BMAD was not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales or transmales **(VERY LOW)**. • Actual femoral area BMAD values were not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transmales or transfemales **(VERY LOW)**. One retrospective observational study (Vlot et al. 2017, n=70) provided non-comparative evidence on change in femoral neck (hip) BMAD between starting GnRH analogues and starting gender-affirming hormones. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. • The z-score for femoral neck BMAD in transfemales with a bone age of <15 years was not statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (z-score median [range]: GnRH analogue −0.71 [−3.35 to 0.37], gender-affirming hormone −1.32 [−3.39 to 0.21], p≤0.1) or in transfemales with a bone age ≥15 years (GnRH analogue −0.44 [−1.37 to 0.93], gender-affirming hormone −0.36 [−1.50 to 0.46]) **(VERY LOW)**. • The z-score for femoral neck BMAD in transmales with a bone age of <14 years was not statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (z-score median [range]: GnRH analogue −0.01

	[−1.30 to 0.91], gender-affirming hormone −0.37 [−2.28 to 0.47]) but was statistically significantly lower in transmales with a bone age ≥14 years (GnRH analogue 0.27 [−1.39 to 1.32], gender-affirming hormone −0.27 [−1.91 to 1.29], p=0.002) **(VERY LOW)**. • Actual femoral neck BMAD values were not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales or in transmales with a young bone age, but were statistically significantly lower in transmales with a bone age ≥14 years (GnRH analogue 0.33 [0.25 to 0.39], gender-affirming hormone 0.30 [0.23 to 0.41], p≤0.01) **(VERY LOW)**. Two uncontrolled, observational, retrospective studies provided evidence for the effect of GnRH analogues on bone density (based on femoral BMD) between starting GnRH analogues and either at 1 or 2 year intervals (Joseph et al. 2019), or starting gender-affirming hormones (Klink et al. 2015). All outcomes were reported separately for transfemales and transmales; also see subgroups table below. One retrospective observational study (Joseph et al. 2019, n=70) provided non-comparative evidence on change in femoral neck BMD increase using z-scores. All outcomes were reported separately for transfemales and transmales. • The z-score for femoral neck BMD was statistically significantly lower at 2 years compared with baseline in transfemales (z-score mean [±SD]: baseline 0.0450 [±0.781], 2 years −0.600 [±1.059], p=0.002) and transmales (baseline −1.075 [±1.145], 2 years −1.779 [±0.816], p=0.001) **(VERY LOW)**. • The z-score for femoral neck BMD was statistically significantly lower at 1 year compared with baseline in transfemales (z-score mean [±SD]: baseline 0.157 [±0.905], 1 year −0.340 [±0.816], p=0.002) and transmales (baseline −0.863 [±1.215], 1 year −1.440 [±1.075], p=0.000) **(VERY LOW)**. • Actual femoral neck BMD values in kg/m² were not statistically significantly different between baseline and 1 or 2 years in transmales or transfemales **(VERY LOW)** One retrospective observational study (Klink et al. 2015, n=34) provided non-comparative evidence on change in femoral area BMD between starting GnRH analogues and starting gender-affirming hormones. All outcomes were reported separately for transfemales and transmales. • The z-score for femoral area BMD was not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales, but was statistically significantly lower in transmales (z-score mean [±SD]: GnRH analogue 0.36 [±0.88], gender-affirming hormone −0.35 [±0.79], p=0.001) **(VERY LOW)**. • Actual femoral area BMD values were not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales, but were statistically significantly lower in transmales (mean [±SD] GnRH analogue 0.92 [±0.10], gender-affirming hormone 0.88 [±0.09], p=0.005) **(VERY LOW)**.

	These studies provide very low certainty evidence that GnRH analogues may reduce the expected increase in femoral bone density (femoral neck or area BMAD or BMD) compared with baseline (although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically significantly decrease actual femoral bone density (femoral area BMAD or femoral neck BMD), apart from actual femoral area BMD in transmales.
Cognitive development or functioning **Certainty of evidence: very low**	This is an important outcome because puberty is an important time for cognitive development and puberty suppression may affect cognitive development or functioning. One cross-sectional observational study (Staphorsius et al. 2015, n=70) provided comparative evidence on cognitive development or functioning in adolescents with gender dysphoria on GnRH analogues compared with adolescents with gender dysphoria not on GnRH analogues. Cognitive functioning was measured using an IQ test. Reaction time (in seconds) and accuracy (percentage of correct trials) were measured using the Tower of London (ToL) task. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. No statistical analyses or interpretation of the results in these groups were reported: • IQ in transfemales (mean [±SD] GnRH analogue 94.0 [±10.3], control 109.4 [±21.2]). IQ transmales (GnRH analogue 95.8 [±15.6], control 98.5 [±15.9]. • Reaction time in transfemales (mean [±SD] GnRH analogue 10.9 [±4.1], control: 9.9 [±3.1]). Reaction time transmales (GnRH analogue 9.9 [±3.1], control 10.0 [±2.0]). • Accuracy score in transfemales (GnRH analogue 73.9 [±9.1], control 83.4 [±9.5]. Accuracy score in transmales (GnRH analogue 85.7 [±10.5], control 88.8 [±9.7]. **This study provides very low certainty evidence (with no statistical analysis) on the effects of GnRH analogues on cognitive development or functioning. No conclusions could be drawn.**
Other safety outcomes: kidney function **Certainty of evidence: very low**	This is an important outcome because if renal damage (raised serum creatinine is a marker of this) is suspected, GnRH analogues may need to be stopped. One prospective observational study (Schagen et al. 2016, n=116) provided non-comparative evidence on change in serum creatinine between starting GnRH analogues and at 1 year. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. • There was no statistically significant difference between baseline and 1 year for serum creatinine in transfemales (mean [±SD] baseline 70 [±12], 1 year 66 [±13], p=0.20). • There was a statistically significant decrease between baseline and 1 year for serum creatinine in transmales (baseline 73 [±8], 1 year 68 [±13], p=0.01). **This study provides very low certainty evidence that GnRH analogues do not affect renal function.**

Other safety outcomes: liver function **Certainty of evidence: very low**	This is an important outcome because if treatment-induced liver injury (raised liver enzymes are a marker of this) is suspected, GnRH analogues may need to be stopped. One prospective observational study (Schagen et al. 2016, n=116) provided non-comparative evidence on elevated liver enzymes between starting GnRH analogues and during use. No comparative values or statistical analyses were reported. • Glutamyl transferase was not elevated at baseline or during use in any person. • Mild elevations of AST and ALT above the reference range were present at baseline but were not more prevalent during use than at baseline. • Glutamyl transferase, AST, and ALT levels did not significantly change from baseline to 12 months of use. **This study provides very low certainty evidence (with no statistical analysis) that GnRH analogues do not affect liver function.**
Other safety outcomes: adverse effects **Certainty of evidence: very low**	This is an important outcome because if there are adverse effects, GnRH analogues may need to be stopped. One uncontrolled, retrospective, observational cohort study (Khatchadourian et al. 2014) provided evidence relating to adverse effects from GnRH analogues. It had incomplete reporting of its cohort, particularly for transfemales where outcomes for only 4/11 were reported. Khatchadourian et al. 2014 reported adverse effects in a cohort of 26 adolescents (15 transmales and 11 transfemales) receiving GnRH analogues. Of these: • 1 transmale developed sterile abscesses; they were switched from leuprolide acetate to triptorelin, and this was well tolerated. • 1 transmale developed leg pains and headaches, which eventually resolved • 1 participant gained 19 kg within 9 months of starting GnRH analogues. **This study provides very low certainty evidence about potential adverse effects of GnRH analogues. No conclusions could be drawn.**

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMAD, bone mineral apparent density; BMD, bone mineral density; GnRH, gonadotrophin releasing hormone; IQ, intelligence quotient; NS, not significant; SD, standard deviation.

In children and adolescents with gender dysphoria, what is the cost-effectiveness of GnRH analogues compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement
Cost-effectiveness	No studies were identified to assess the cost-effectiveness of GnRH analogues for children and adolescents with gender dysphoria.

From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may benefit from GnRH analogues more than the wider population of interest?

Subgroup	Evidence statement
Sex assigned at birth males (transfemales) **Certainty of evidence: Very low**	Some studies reported data separately for sex assigned at birth males (transfemales). This included some direct comparisons with sex assigned at birth females (transmales). **Impact on gender dysphoria** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence for gender dysphoria in sex assigned at birth males. See the clinical effectiveness results table above for a full description of the study. The mean (±SD) UGDS score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean UGDS score [±SD]: 47.95 [±9.70] versus 56.57 [±3.89]) and T1 (n=not reported, 49.67 [±9.47] versus 56.62 [±4.00]), between sex difference p<0.001 **(VERY LOW)**. One further prospective observational longitudinal study (Costa et al. 2015) provided evidence for the impact on gender dysphoria in sex assigned at birth males. See the clinical effectiveness results table above for a full description of the study. Sex assigned at birth males had a statistically significantly lower (improved) mean (±SD) UGDS score of 51.6 [±9.7] compared with sex assigned at birth females (56.1 [±4.3], p<0.001). However, it was not reported if this was baseline or follow-up **(VERY LOW)**. **These studies provide very low certainty evidence that in sex assigned at birth males (transfemales), gender dysphoria is lower than in sex assigned at birth females (transmales).** **Impact on mental health** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence for the impact on mental health (depression, anger and anxiety) in sex assigned at birth males. See the clinical effectiveness results table above for a full description of the study. • The mean (±SD) depression (BDI-II) score was not statistically significantly different in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean BDI score [±SD]: 5.71 [±4.31] versus 10.34 [±8.24]) and T1 (n=not reported, 3.50 [±4.58] versus 6.09 [±7.93]), between sex difference p=0.057 • The mean (±SD) anger (TPI) score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean TPI score [±SD]: 5.22 [±2.76] versus 6.43 [±2.78]) and T1 (n=not reported, 5.00 [±3.07] versus 6.39 [±2.59]), between sex difference p=0.022 • The mean (±SD) anxiety (STAI) score was statistically significantly lower (improved) in sex assigned at birth males

compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean STAI score [±SD]: 4.33 [±2.68] versus 7.00 [±2.36]) and T1 (n=not reported, 4.39 [±2.64] versus 6.17 [±2.69]), between sex difference p<0.001 **(VERY LOW)**.

This study provides very low certainty evidence that the impact on mental health (depression, anger and anxiety) may be different in sex assigned at birth males (transfemales) compared with sex assigned at birth females (transmales). Over time there was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for depression. However, sex assigned at birth males had statistically significantly lower levels of anger and anxiety than sex assigned at birth females at both baseline and follow up.

Impact on body image

One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on body image in sex assigned at birth males.

- The mean (±SD) BIS score for primary sex characteristics was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean BIS score [±SD]: 4.02 [±0.61] versus 4.16 [±0.52]) and T1 (n=not reported, 3.74 [±0.78] versus 4.17 [±0.58]), between sex difference p=0.047
- The mean (±SD) BIS score for secondary sex was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean BIS score [±SD]: 2.66 [±0.50] versus 2.81 [±0.76]) and T1 (n=not reported, 2.39 [±0.69] versus 3.18 [±0.42]), between sex difference p=0.001
- The mean (±SD) BIS score for neutral body characteristics was not statistically significantly different in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean BIS score [±SD]: 2.60 [±0.58] versus 2.24 [±0.62]) and T1 (n=not reported, 2.32 [±0.59] versus 2.61 [±0.50]), between sex difference p=0.777 **(VERY LOW)**.

This study provides very low certainty evidence that the impact on body image may be different in sex assigned at birth males (transfemales) compared with sex assigned at birth females (transmales). Sex assigned at birth males are less dissatisfied with their primary and secondary sex characteristics than sex assigned at birth females at both baseline and follow up, but the satisfaction with neutral body characteristics is not different.

Psychosocial impact

One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence for psychosocial impact in terms of global functioning (CGAS) and psychosocial functioning (CBCL and YSR) in sex assigned at birth males.

- Sex assigned at birth males had statistically higher mean (±SD) CGAS scores compared with sex assigned at birth

	females at both baseline (T0) (n=54, 73.10 [±8.44] versus 67.25 [±11.06]) and T1 (n=54, 77.33 [±8.69] versus 70.30 [±9.44]), between sex difference p=0.021 • There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for the CBCL Total T score at T0 or T1 (n=54, p=0.110) • There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for the CBCL internalising T score at T0 or T1 (n=54, p=0.286) • Sex assigned at birth males had statistically lower mean (±SD) CBCL externalising T scores compared with sex assigned at birth females at both T0 (n=54, 54.71 [±12.91] versus 60.70 [±12.64]) and T1 (n=54, 48.75 [±10.22] versus 57.87 [±11.66]), between sex difference p=0.015 • There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for the YSR Total T score at T0 or T1 (n=54, p=0.164) • There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for the YSR internalising T score at T0 or T1 (n=54, p=0.825) • Sex assigned at birth males had statistically lower mean (±SD) YSR externalising T scores compared with sex assigned at birth females at both T0 (n=54, 48.72 [±11.38] versus 57.24 [±10.59]) and T1 (n=54, 46.52 [±9.23] versus 52.97 [±8.51]), between sex difference p=0.004 **(VERY LOW)**. One uncontrolled, observational, prospective cohort study (Costa et al. 2015) provided evidence for psychosocial impact in terms of global functioning (CGAS) in sex assigned at birth males. • Sex assigned at birth males had statistically significant lower mean (±SD CGAS scores at baseline) compared with sex assigned at birth females (n=201, 55.4 [±12.7] versus 59.2 [±11.8], p=0.03) **(VERY LOW)**. **These studies provide very low certainty evidence that psychosocial impact may be different in sex assigned at birth males (transfemales) compared with sex assigned at birth females (transmales). However, no conclusions could be drawn.** **Change in bone density: lumbar** Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on lumbar bone density in sex assigned at birth males (Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results. **These studies provide very low certainty evidence that GnRH analogues reduce the expected increase in lumbar bone density (BMAD or BMD) in sex assigned at birth males (transfemales; although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically significantly decrease actual lumbar bone density (BMAD or BMD) in sex assigned at birth males (transfemales).** **Change in bone density: femoral**

	Three uncontrolled, observational, retrospective studies provided evidence for the effect of GnRH analogues on femoral bone density in sex assigned at birth males (Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results. **These studies provide very low certainty evidence that GnRH analogues may reduce the expected increase in femoral bone density (femoral neck or area BMAD or BMD) in sex assigned at birth males (transfemales; although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically significantly decrease actual femoral bone density (femoral area BMAD or femoral neck BMD) in sex assigned at birth males (transfemales).** **Cognitive development or functioning** One cross-sectional observational study (Staphorsius et al. 2015) provided comparative evidence on cognitive development or functioning in sex assigned at birth males. See the safety results table above for a full description of the results. **This study provides very low certainty evidence (with no statistical analysis) on the effects of GnRH analogues on cognitive development or functioning in sex assigned at birth males (transfemales). No conclusions could be drawn.** **Other safety outcomes: kidney function** One prospective observational study (Schagen et al. 2016) provided non-comparative evidence on change in serum creatinine in sex assigned at birth males. See the safety results table above for a full description of the results. **This study provides very low certainty evidence that GnRH analogues do not affect renal function in sex assigned at birth males (transfemales).**
Sex assigned at birth females (transmales) **Certainty of evidence: Very low**	Some studies reported data separately for sex assigned at birth females (transmales). This included some direct comparisons with sex assigned at birth males (transfemales). **Impact on gender dysphoria** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) and one prospective observational longitudinal study (Costa et al. 2015) provided evidence for gender dysphoria in sex assigned at birth females. See the sex assigned at birth males (transfemales) row above for a full description of the results. **These studies provide very low certainty evidence that in sex assigned at birth females (transmales), gender dysphoria is higher than in sex assigned at birth males (transfemales) at both baseline and follow up.** **Impact on mental health** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on mental health (depression, anger and anxiety) in sex assigned at birth

	females. See the sex assigned at birth males (transfemales) row above for a full description of the results. **This study provides very low certainty evidence that the impact on mental health (depression, anger and anxiety) may be different in sex assigned at birth females (transmales) compared with sex assigned at birth males (transfemales). Over time there was no statistically significant difference between sex assigned at birth females and sex assigned at birth males for depression. However, sex assigned at birth females had statistically significantly greater levels of anger and anxiety than sex assigned at birth males at baseline and follow up.** **Impact on body image** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on body image in sex assigned at birth females. See the sex assigned at birth males (transfemales) row above for a full description of the results. **This study provides very low certainty evidence that the impact on body image may be different in sex assigned at birth females (transmales) compared with sex assigned at birth males (transfemales). Sex assigned at birth females are more dissatisfied with their primary and secondary sex characteristics than sex assigned at birth males at both baseline and follow up, but the satisfaction with neutral body characteristics is not different.** **Psychosocial impact** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence for psychosocial impact in terms of global functioning (CGAS) and psychosocial functioning (CBCL and YSR) in sex assigned at birth females. One uncontrolled, observational, prospective cohort study (Costa et al. 2015) provided evidence for psychosocial impact in terms of global functioning (CGAS) in sex assigned at birth females. See the sex assigned at birth males (transfemales) row above for a full description of the results. **These studies provide very low certainty evidence that psychosocial impact may be different in sex assigned at birth females (transmales) compared with sex assigned at birth males (transfemales). However, no conclusions could be drawn.** **Change in bone density: lumbar** Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on lumbar bone density in sex assigned at birth females (Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results. **These studies provide very low certainty evidence that GnRH analogues reduce the expected increase in lumbar bone density (BMAD or BMD) in sex assigned at birth females (transmales; although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically**

	significantly decrease actual lumbar bone density (BMAD or BMD) in sex assigned at birth females (transmales). **Change in bone density: femoral** Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on femoral bone density in sex assigned at birth females (Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results. **These studies provide very low certainty evidence that GnRH analogues may reduce the expected increase in femoral bone density (femoral neck or area BMAD or BMD) in sex assigned at birth females (transmales; although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically significantly decrease actual femoral bone density (femoral area BMAD or femoral neck BMD) in sex assigned at birth females (transmales), apart from actual femoral area.** **Cognitive development or functioning** One cross-sectional observational study (Staphorsius et al. 2015) provided comparative evidence on cognitive development or functioning in sex assigned at birth females. See the safety results table above for a full description of the results. **This study provides very low certainty evidence (with no statistical analysis) on the effects of GnRH analogues on cognitive development or functioning in sex assigned at birth females (transmales). No conclusions could be drawn.** **Other safety outcomes: kidney function** One prospective observational study (Schagen et al. 2016) provided non-comparative evidence on change in serum creatinine in sex assigned at birth females (transmales). See the safety results table above for a full description of the results. **This study provides very low certainty evidence that GnRH analogues do not affect renal function in sex assigned at birth females (transmales).**
Duration of gender dysphoria	No evidence was identified.
Age at onset of gender dysphoria	No evidence was identified.
Age at which GnRH analogue started	No evidence was identified.
Age at onset of puberty	No evidence was identified.
Tanner stage at which GnRH analogue started	No evidence was identified.
Diagnosis of autistic spectrum disorder	No evidence was identified.

Diagnosis of mental health condition	No evidence was identified.

Abbreviations: BDI-II, Beck Depression Inventory-II; BIS, Body Image Scale; CBCL, Child Behaviour Checklist; CGAS, Children's Global Assessment Scale; SD, standard deviation; STAI, Trait Anxiety Scale of the State-Trait Personality Inventory; TPI, Trait Anger Scale of the State-Trait Personality Inventory; UGDS, Utrecht Gender Dysphoria Scale; YSR, Youth Self-Report

From the evidence selected,

(a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?

(b) what were the ages at which participants commenced treatment with GnRH analogues?

(c) what was the duration of treatment with GnRH analogues?

<table>
<tr><th>Outcome</th><th>Evidence statement</th></tr>
<tr><td>Diagnostic criteria</td><td>In 5 studies (Costa et al. 2015, Klink et al. 2015, Schagen et al. 2016, Staphorsius et al. 2015 and Vlot et al. 2017) the DSM-IV-TR criteria of gender identity disorder was used.

The study by Brik et al. 2020 used DSM-V criteria. The DSM-V has one overarching definition of gender dysphoria with separate specific criteria for children and for adolescents and adults. The general definition describes a conflict associated with significant distress and/or problems functioning associated with this conflict between the way they feel and the way they think of themselves which must have lasted at least 6 months.

It was not reported how gender dysphoria was defined in the remaining 3 studies (VERY LOW).

From the evidence selected, all studies that reported diagnostic criteria for gender dysphoria (6/9 studies) used the DSM criteria in use at the time the study was conducted.</td></tr>
<tr><td>Age when GnRH analogues started</td><td>8/9 studies reported the age at which participants started GnRH analogues, either as the mean age (with SD) or median age (with the range):

<table>
<tr><th>Study</th><th>Mean age (±SD)</th></tr>
<tr><td>Costa et al. 2015</td><td>16.5 years (±1.3)</td></tr>
<tr><td>de Vries et al. 2011</td><td>13.6 years (±1.8)</td></tr>
<tr><td>Joseph et al. 2019</td><td>13.2 years (±1.4) in transfemales
12.6 years (±1.0) in transmales</td></tr>
<tr><td>Khatchadourian et al. 2014</td><td>14.7 years (±1.9)</td></tr>
<tr><td>Klink et al. 2015</td><td>14.9 years (±1.9) in transfemales
15.0 years (±2.0) in transmales</td></tr>
</table>
<table>
<tr><th>Study</th><th>Median age (range)</th></tr>
<tr><td>Brik et al. 2020</td><td>15.5 years (11.1–18.6) in transfemales
16.1 years (10.1–17.9) in transmales</td></tr>
</table>
</td></tr>
</table>

	Schagen et al. 2016	13.6 years (11.6–17.9) in transfemales 14.2 years (11.1–18.6) in transmales
	Vlot et al. 2017	13.5 years (11.5–18.3) in transfemales 15.1 years (11.7–18.6) in transmales
	Age at the start of GnRH analogues was not reported in Staphorsius et al. 2015, but participants were required to be at least 12 years **(VERY LOW)**. **The evidence included showed wide variation in the age (11 to 18 years old) at which children and adolescents with gender dysphoria started GnRH analogues.**	
Duration of treatment	The duration of treatment with GnRH analogues was reported in 3/9 studies. The median duration was: • 2.1 years (range 1.6–2.8) in Brik et al. 2020. • 1.3 years (range 0.5–3.8) in transfemales and 1.5 years (range 0.25–5.2) in transmales in Klink et al. 2015. In Staphorsius et al. 2015, the mean duration was 1.6 years (SD ±1.0). In de Vries et al. 2011, the mean duration of time between starting GnRH analogues and gender-affirming hormones was 1.88 years (SD ±1.05). **The evidence included showed wide variation in the duration of treatment with GnRH analogues, but most studies did not report this information. Treatment duration ranged from a few months up to about 5 years.**	

Abbreviations: DSM, Diagnostic and Statistical Manual of Mental Disorders criteria; SD, standard deviation.

6. Discussion

A key limitation to identifying the effectiveness and safety of GnRH analogues for children and adolescents with gender dysphoria is the lack of reliable comparative studies. The lack of clear, expected outcomes from treatment with a GnRH analogue (the purpose of which is to suppress secondary sexual characteristics which may cause distress from unwanted pubertal changes) also makes interpreting the evidence difficult. The size of the population with gender dysphoria means conducting a prospective trial may be unrealistic, at least on a single centre basis. There may also be ethical issues with a 'no treatment arm' in comparative trials of GnRH analogues, where there may be poor mental health outcomes if treatment is withheld. However, the use of an active comparator such as close psychological support may reduce ethical concerns in future trials.

The studies included in this evidence review are all small, uncontrolled observational studies, which are subject to bias and confounding, and are of very low certainty as assessed using modified GRADE. All the included studies reported physical and mental health comorbidities and concomitant treatments very poorly. For example, very little data are reported on how many children and adolescents needed additional mental health support, and for what reasons, or whether additional interventions, and what form and duration (for example drug treatment or counselling) that took. This is a possible confounder for the treatment outcomes in the studies because changes in critical and important

outcomes may be attributable to external care rather than the psychological support or GnRH analogues used in the studies.

The studies that reported diagnostic criteria for gender dysphoria (6/9 studies) used the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria in use at the time the study was conducted (either DSM-IV-TR or DSM-V). The definition was unclear in the remaining studies. There was wide variation in the ages at which participants started a GnRH analogue, typically ranging from about 11 to 18 years. Similarly, there was a wide variation in the duration of use, but few studies reported this.

Changes in outcome scores for clinical effectiveness were assessed for statistical significance in the 3 studies reporting these outcomes (Costa et al. 2015; de Vries et al. 2011; Staphorsius et al. 2015). However, there is relatively little interpretation of whether the changes in outcome scores seen in these studies are clinically meaningful.

For some outcomes there was no statistically significant difference from before starting GnRH analogues until just before starting gender-affirming hormones. These were the Utrecht Gender Dysphoria Scale (UGDS) (which was assessed in 1 study de Vries et al. 2011), the Trait Anger (TPI) and Trait Anxiety (STAI) Scales (which were assessed in 1 study de Vries et al. 2011), and Body Image Scale (BIS) which was assessed in 1 study (de Vries et al. 2011).

The Beck Depression Inventory (BDI-II) was used in 1 study (de Vries et al. 2011) to assess change in depression from before starting GnRH analogues to just before starting gender-affirming hormones. The result is statistically significant, with the mean (±SD) BDI-II score decreasing from 8.31 (±7.12) at baseline to 4.95 (±6.27) at follow up (p=0.004). However, both scores fall into the minimal range using the general guidelines for interpretation of BDI-II (0 to 13 minimal, 14 to 19 mild depression, 20 to 28 moderate depression and 29 to 63 severe depression), suggesting that while statistically significant, it is unclear if this is a clinically meaningful change.

Psychosocial outcomes were assessed in 3 studies (Costa et al. 2015; de Vries et al. 2011; Staphorsius et al. 2015) using the Children's Global Assessment Scale (CGAS) and Child Behavior Checklist/Youth Self-Report (CBCL/YSR). The CGAS score was assessed in 2 studies (Costa et al. 2015; de Vries et al. 2011). In de Vries et al. 2011 the mean (±SD) CGAS score statistically significantly increased over time from 70.24 [±10.12] at baseline to 73.90 [±9.63] at follow up. CGAS scores are clinically categorised into 10 categories (10 to 1, 20 to 11 and so on until 100 to 91) and both scores reported were in a single category (71 to 80, no more than slight impairment) suggesting that while statistically significant, it is unclear if this is a clinically meaningful change. The Costa et al. 2015 study does highlight a larger change in CGAS scores from baseline to follow-up (mean [±SD] 58.72 [±11.38] compared with 67.40 [±13.39]), but whether this is clinically meaningful is unclear. The average score moved from the clinical category of 60 to 51 (variable functioning with sporadic difficulties) at baseline to 70 to 61 (some difficulty in a single area, but generally functioning pretty well) at follow up, but the large standard deviations suggest clinically significant overlaps between the scores from baseline to follow-up.

Psychosocial functioning using the CBCL/YSR was assessed in 2 studies (de Vries et al. 2011; Staphorsius et al. 2015). In de Vries et al. 2011 there was a statistically significant reduction in both CBCL and YSR scores from before starting GnRH analogues to just before

starting gender-affirming hormones. The study interpreted the CBCL/YSR with a proportion of adolescents who scored in the clinical range (a T-score above 63), which allows changes in clinically meaningful scores to be assessed, and proportions of adolescents in the clinical range for some CBCL and YSR scores decreased over time. One cross-sectional study (Staphorsius et al. 2015) assessed CBCL scores only, but it was unclear if this was the Total T score, or whether subscales of internalising or externalising scores were also assessed, and whether the results were statistically significant.

The 2 prospective observational studies (Costa et al. 2015; de Vries et al. 2011) are confounded by a number of common factors. Firstly, the single assessment of scores at baseline means it is unclear if scores were stable, already improving or declining before starting treatment. Secondly, in an uncontrolled study any changes in scores from baseline to follow-up could be attributed to a regression-to-mean, for example getting older has been positively associated with maturity and wellbeing. The studies use mean and standard deviations in the descriptive statistics and analyses; however, they do not report testing the normality of data which would support the use of parametric measures. The study by de Vries et al. 2011 used general linear models (regression) to examine between and within group variances (changes in outcomes). In using such models, the data is assumed to be balanced (measured at regular intervals and without missing data), but the large ranges in ages at which participants were assessed and started on various interventions suggests that ascertainment of outcome was unlikely to be regular and missing data was likely. Missing data was handled through listwise deletion (omits those cases with the missing data and analyses the remaining data) which is acceptable if data loss is completely random but for some outcomes where there was incomplete data for individual items this was not random (items were introduced by the authors after the first eligible adolescents had started GnRH analogues). The study provided no detail on whether these assumptions for the modeling were met, they also provided no adequate assessment of whether any regression diagnostics (analysis that seek to assess the validity of a model) or model fit (how much of the variance in outcome is explained by the between and within group variance) were undertaken.

The 2 retrospective observational studies (Brik et al. 2020; Khatchadourian et al. 2014) both only report absolute numbers for each trajectory along with reasons for stopping GnRH analogues. It is difficult to assess outcomes from such single centre studies because there is little comparative data for outcomes from other such services. A lack of any critical or other important outcomes also means the success of the treatment across all the participants is difficult to judge.

Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on bone density (Joseph et al. 2019; Klink et al. 2015; Vlot et al. 2017). In all 3 studies, the participants acted as their own controls and change in bone density was determined between starting GnRH analogues and either after 1 and 2 year follow-up timepoints (Joseph et al. 2019) or when gender-affirming hormones were started (Klink et al. 2015 and Vlot et al. 2017). Observational studies such as these can only show an association with GnRH analogues and bone density; they cannot show that GnRH analogues caused any differences in bone density seen. Because there was no comparator group and participants acted as their own controls, it is unclear whether the findings are associated with GnRH analogues or due to changes over time. The authors reported z-scores which allows for comparison with the expected increase in bone density in the

general population. However, because no concomitant treatments or comorbidities were reported it is possible that the findings may not be because of GnRH analogues and there is another way in which the study population differs from the general population.

All the studies are from a limited number of, mainly European, care facilities. They are described as either tertiary referral or expert services but the low number of services providing such care and publishing evidence may bias the results towards the outcomes in these services only and limit extrapolation.

The first study (Brik et al. 2020) was an uncontrolled, retrospective, observational study that assessed the outcome trajectories of adolescents receiving GnRH analogues for gender dysphoria. This study followed-up 143 individuals who had received GnRH analogues (38 transfemales and 105 transmales) using clinical records to show outcomes for up to 9 years (continuing use of GnRH analogues, reasons for stopping GnRH analogues and onward care such as gender-affirming hormone use). The methods and results are well reported, but no analysis of data was undertaken. The views of adolescents and their parents are particularly difficult to interpret because no data on how many responded to each question and in what ways are reported.

The second study (Costa et al. 2015) was an uncontrolled, prospective observational study which assessed global functioning in adolescents with gender dysphoria using CGAS every 6 months, including during the first 6 months where statistically significant improvements were seen without GnRH analogues. The study is confounded by significant unexplained loss to follow-up (64.7% from n=201 adolescents to n=71 after 18 months). Missing data for those lost to follow-up maybe more than sufficient to change the direction of effects seen in the study if the reasons for loss to follow-up are systematic (such as deriving little or no benefit from treatment). The study uses clustered data in its analysis, a single outcome (CGAS) measured in clusters (at different visits), and the analysis does not take account of the correlation of scores (data at different time points are not independent) as a significant change in scores early in the study means the successive changes measured against baseline were also significant. The study relies on multiple (>20) pairwise independent *t*-tests to examine change in CGAS between the 4 time points, increasing the possibility of type-I error (a false positive which occurs when a researcher incorrectly rejects a true null hypothesis) because the more tests performed the more likely a statistically significant result will be observed by chance alone.

The Costa et al. 2015 study compares immediately eligible and delayed eligible cohorts, however, it is highly likely that they are non-comparable groups because the immediately eligible group were those able to start GnRH analogues straight away whilst those in the delayed eligible group were either not ready to make a decision about starting treatment (no age comparison was made between the 2 groups so it is unclear if they were a younger cohort than the immediately eligible group) or had comorbid mental health or psychological difficulties. The authors report that those with concomitant problems (such as mental health problems, substantial problems with peers, or conflicts with parents or siblings) were referred to local mental health services but no details are provided.

The third study (de Vries et al. 2011) was an uncontrolled, prospective observational study which assessed gender dysphoria and psychological functioning before and after puberty suppression in adolescents with gender dysphoria. Although the study mentions the DSM-IV-TR there is no explicit discussion of this, or any other criteria, being used as the

diagnostic criteria for study entry. There are no details reported for how the outcomes in the study were assessed, and by whom. The length of follow-up for the outcomes in the model are questionable in relation to whether there was sufficient time for GnRH analogues to have a measurable effect. The time points used are start of GnRH analogues and start of gender-affirming hormones. Overall, the mean time between starting GnRH analogues and gender-affirming hormones was 1.88 (±1.05) years, but the range is as low as just 5 months between the 2 time points, which may be insufficient for any difference in outcome to have occurred in some individuals.

The fourth study (Joseph et al. 2019) was a retrospective, longitudinal observational single centre study which assessed bone mineral density in adolescents with gender dysphoria in the UK. For inclusion in the study, participants had to have been assessed by the Gender Identity Development Service multi-disciplinary psychosocial health team for at least 4 assessments over a minimum of 6 months. No other diagnostic criteria, such as the DSM-IV-TR, are discussed. Bone density was assessed using dual energy X-ray absorptiometry (DAXA) scan of the lumbar spine (L1-L4) and the femoral neck at baseline (n=70), 1 year (n=70) and 2 years after starting GnRH analogues (n=39). The results suggest a possible association between GnRH analogues and bone mineral apparent density. However, the evidence is of poor quality, and the results could be due to bias or chance. No concomitant treatments or comorbidities were reported.

The fifth study (Khatchadourian et al. 2014) was an uncontrolled retrospective observational study which describes patient characteristics at presentation, treatment, and response to treatment in 84 adolescents with gender dysphoria, of whom 27 received GnRH analogues. The study used clinical records to show outcomes for up to 13 years (continuing use of GnRH analogues, reasons for stopping GnRH analogues and onward care such as gender-affirming hormone use). The methods are well reported but the results for those taking GnRH analogues are poorly and incompletely reported, particularly for transfemales, and no analysis of data was undertaken. It is difficult to assess the results for stopping GnRH analogues due to incomplete reporting of this outcome.

The sixth study (Klink et al. 2015) was a retrospective longitudinal observational single centre study which assessed bone mineral density in adolescents with gender dysphoria, diagnosed with the DSM-IV-TR criteria. Bone density was assessed when starting GnRH analogues and then when starting gender-affirming hormones. Results are reported for transmales and transfemales separately and no results for the whole cohort are given. Statistical analyses were reported for all outcomes of interest but, because there was no comparator group and participants acted as their own controls, it is not known whether the findings are associated with GnRH analogues or due to changes over time. The authors reported z-scores which allows for comparison with the expected increase in bone density in the general population. However, because no concomitant treatments or comorbidities were reported it is possible that the findings may not be because of GnRH analogues and there is another way in which the study population differs from the general population.

The seventh study (Schagen et al. 2016) was a prospective observational study of 116 adolescents which provided very low certainty non-comparative evidence on change in serum creatinine between starting GnRH analogues and 1 year, and liver function during treatment. Statistical analyses were reported for changes in serum creatinine but not for liver function. Because there was no comparator group and participants acted as their own

controls, it is not known whether the findings are associated with GnRH analogues or due to changes over time, or concomitant treatments.

The eighth study (Staphorsius et al. 2015) was a cross-sectional study of 85 adolescents, 40 with gender dysphoria (of whom 20 were receiving GnRH analogues) and 45 matched controls (not further reported in this evidence review). The study includes 1 outcome of interest for clinical effectiveness (CBCL) and 1 outcome of interest for safety (cognitive development or functioning). The mean (±SD) CBCL, IQ test, reaction time and accuracy scores were given for each group, but the statistical analysis is unclear. It is not reported what analysis was used or which of the groups were compared, therefore it is difficult to interpret the results.

The ninth study (Vlot et al. 2017) was a retrospective observational study which assessed bone mineral apparent density in adolescents with DSM-IV-TR gender dysphoria. Measurements were taken at the start of GnRH analogues and at the start of gender-affirming hormones. Results are reported for young bone age and old bone age in transmales and transfemales separately, and no results for the whole cohort are given. Statistical analyses were reported for all outcomes of interest but, because there was no comparator group and participants acted as their own controls, it is not known whether the findings are associated with GnRH analogues or due to changes over time. The authors reported z-scores which allows for comparison with the expected increase in bone density in the general population. However, because no concomitant treatments or comorbidities were reported it is possible that the findings may not be because of GnRH analogues and there is another way in which the study population differs from the general population.

7. Conclusion

The results of the studies that reported impact on the critical outcomes of gender dysphoria and mental health (depression, anger and anxiety), and the important outcomes of body image and psychosocial impact (global and psychosocial functioning) in children and adolescents with gender dysphoria are of very low certainty using modified GRADE. They suggest little change with GnRH analogues from baseline to follow-up.

Studies that found differences in outcomes could represent changes that are either of questionable clinical value, or the studies themselves are not reliable and changes could be due to confounding, bias or chance. It is plausible, however, that a lack of difference in scores from baseline to follow-up is the effect of GnRH analogues in children and adolescents with gender dysphoria, in whom the development of secondary sexual characteristics might be expected to be associated with an increased impact on gender dysphoria, depression, anxiety, anger and distress over time without treatment. One study reported statistically significant reductions in the Child Behaviour Checklist/Youth Self-Report (CBCL/YSR) scores from baseline to follow up, and given that the purpose of GnRH analogues is to reduce distress caused by the development of secondary sexual characteristics and the CBCL/YSR in part measures distress, this could be an important finding. However, as the studies all lack reasonable controls not receiving GnRH analogues, the natural history of the outcomes measured in the studies is not known and any positive changes could be a regression to mean.

The results of the studies that reported bone density outcomes suggest that GnRH analogues may reduce the increase in bone density which is expected during puberty. However, as the studies themselves are not reliable, the results could be due to confounding, bias or chance. While controlled trials may not be possible, comparative studies are needed to understand this association and whether the effects of GnRH analogues on bone density are seen after treatment is stopped. All the studies that reported safety outcomes provided very low certainty evidence.

No cost-effectiveness evidence was found to determine whether or not GnRH analogues are cost-effective for children and adolescents with gender dysphoria.

The results of the studies that reported outcomes for subgroups of children and adolescents with gender dysphoria, suggest there may be differences between sex assigned at birth males (transfemales) and sex assigned at birth females (transmales).

Appendix A PICO document

The review questions for this evidence review are:

1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
2. For children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
3. For children and adolescents with gender dysphoria, what is the cost-effectiveness of GnRH analogues compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?
4. From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may derive more (or less) advantage from treatment with GnRH analogues than the wider population of children and adolescents with gender dysphoria?
5. From the evidence selected,
   a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?
   b) what were the ages at which participants commenced treatment with GnRH analogues?
   c) what was the duration of treatment with GnRH analogues?

PICO table

P – Population and Indication	Children and adolescents aged 18 years or less who have gender dysphoria, gender identity disorder or gender incongruence of childhood as defined by study: The following subgroups of children and adolescents with gender dysphoria, gender identity disorder or gender incongruence of childhood need to be considered: • Sex assigned at birth males. • Sex assigned at birth females. • The duration of gender dysphoria: less than 6 months, 6-24 months, and more than 24 months. • The age of onset of gender dysphoria. • The age at which treatment was initiated. • The age of onset of puberty. • Tanner stage at which treatment was initiated. • Children and adolescents with gender dysphoria who have a pre-existing diagnosis of autistic spectrum disorder. • Children and adolescents with gender dysphoria who had a significant mental health symptom load at diagnosis including anxiety, depression (with or without a history of self-harm and suicidality), suicide attempts, psychosis, personality disorder, Attention Deficit Hyperactivity Disorder and eating disorders.
I – Intervention	Any GnRH analogue including: triptorelin*; buserelin; histrelin; goserelin (Zoladex); leuprorelin/leuprolide (Prostap); nafarelin.

	* Triptorelin (brand names Gonapeptyl and Decapeptyl) are used in Leeds Hospital, England. The search should include brand names as well as generic names.
C – Comparator(s)	One or a combination of: • Psychological support. • Social transitioning to the gender with which the individual identifies. • No intervention.
O – Outcomes	There are no known minimal clinically important differences and there are no preferred timepoints for the outcome measures selected. **All outcomes should be stratified by:** • The age at which treatment with GnRH analogues was initiated. • The length of treatment with GnRH analogues where possible. **A: Clinical Effectiveness** *Critical to decision making* • **Impact on Gender Dysphoria** This outcome is critical because gender dysphoria in adolescents and children is associated with significant distress and problems functioning. Impact on gender dysphoria may be measured by the Utrecht Gender Dysphoria Scale. Other measures as reported in studies may be used as an alternative to the stated measure. • **Impact on mental health** Examples of mental health problems include self-harm, thoughts of suicide, suicide attempts, eating disorders, depression/low mood and anxiety. These outcomes are critical because self-harm and thoughts of suicide have the potential to result in significant physical harm and for completed suicides the death of the young person. Disordered eating habits may cause significant morbidity in young people. Depression and anxiety are also critical outcomes because they may impact on social, occupational, or other areas of functioning of children and adolescents. The Child and Adolescent Psychiatric Assessment (CAPA) may be used to measure depression and anxiety. The impact on self-harm and suicidality (ideation and behaviour) may be measured using the Suicide Ideation Questionnaire Junior. Other measures may be used as an alternative to the stated measures. • **Impact on Quality of Life** This outcome is critical because gender dysphoria in children and adolescents may be associated with a significant reduction in health-related quality of life. Quality of Life may be measured by the KINDL questionnaire, Kidscreen 52. Other measures as reported in studies may be used as an alternative to the stated measure. *Important to decision making* • **Impact on body Image** This outcome is important because some transgender young people may desire to take steps to suppress features of their physical appearance associated with their sex assigned at birth or accentuate physical features of their desired gender. The Body Image Scale could be used as a measure. Other measures

	as reported in studies may also be used as an alternative to the stated measure. • **Psychosocial Impact** Examples of psychosocial impact are: coping mechanisms which may impact on substance misuse; family relationships; peer relationships. This outcome is important because gender dysphoria in adolescents and children is associated with internalising and externalising behaviours and emotional and behavioural problems which may impact on social and occupational functioning. The child behavioural check list (CBCL) may be used to measure the impact on psychosocial functioning. Other measures as reported in studies may be used as an alternative to the stated measure. • **Engagement with health care services** This outcome is important because patient engagement with healthcare services will impact on their clinical outcomes. Engagement with health care services may be measured using the Youth Health Care measure-satisfaction, utilization, and needs (YHC-SUN) questionnaire. Loss to follow up should also be ascertained as part of this outcome. Alternative measures to the YHC-SUN questionnaire may be used as reported in studies. • **Transitioning surgery – Impact on extent of and satisfaction with surgery** This outcome is important because some children and adolescents with gender dysphoria may proceed to transitioning surgery. Stated measures of the extent of transitioning surgery and satisfaction with surgery in studies may be reported. • **Stopping treatment** The proportion of patients who stop treatment with GnRH analogues and the reasons why. This outcome is important to patients because there is uncertainty about the short- and long-term safety and adverse effects of GnRH analogues in children and adolescents being treated for gender dysphoria. **B: Safety** • Short and long-term safety and adverse effects of taking GnRH analogues are important because GnRH analogues are not licensed for the treatment of adolescents and children with gender dysphoria. Aspects to be reported on should include: ○ Impact of the drug use such as its impact on bone density, arterial hypertension, cognitive development/functioning ○ Impact of withdrawing the drug such as, slipped upper femoral epiphysis, reversibility on the reproductive system, and any others as reported. **C: Cost effectiveness** Cost effectiveness studies should be reported.
Inclusion criteria	
Study design	Systematic reviews, randomised controlled trials, controlled clinical trials, cohort studies. If no higher level quality evidence is found, case series can be considered.

Language	English only
Patients	Human studies only
Age	18 years or less
Date limits	2000-2020
Exclusion criteria	
Publication type	Conference abstracts, non-systematic reviews, narrative reviews, commentaries, letters, editorials, guidelines and pre-publication prints
Study design	Case reports, resource utilisation studies

Appendix B Search strategy

Medline, Embase, the Cochrane Library, HTA and APA PsycInfo were searched on 23 July 2020, limiting the search to papers published in English language in the last 20 years. Conference abstracts and letters were excluded.

Database: Medline
Platform: Ovid
Version: Ovid MEDLINE(R) <1946 to July 21, 2020>
Search date: 23/7/2020
Number of results retrieved: 144
Search strategy:

1 Gender Dysphoria/ (485)
2 Gender Identity/ (18452)
3 "Sexual and Gender Disorders"/ (75)
4 Transsexualism/ (3758)
5 Transgender Persons/ (3143)
6 Health Services for Transgender Persons/ (136)
7 exp Sex Reassignment Procedures/ (836)
8 (gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (7435)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (12678)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (102343)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (6974)
12 (male-to-female or m2f or female-to-male or f2m).tw. (114841)
13 or/1-12 (252702)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (1137479)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (852400)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (1913257)

17 Minors/ (2574)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (2361686)
19 exp pediatrics/ (58118)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (836269)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (2024207)
22 Puberty/ (13278)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (424246)
24 Schools/ (38104)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (7199)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (468992)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (89353)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (887838)
29 or/14-28 (5534171)
30 13 and 29 (79263)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (7)
32 30 or 31 (79263)
33 Gonadotropin-Releasing Hormone/ (27588)
34 (pubert* adj3 block*).ti,ab. (78)
35 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (17299)
36 (GnRH adj2 analog*).ti,ab. (2541)
37 GnRH*.ti,ab. (20991)
38 "GnRH agonist*".ti,ab. (4040)
39 Triptorelin Pamoate/ (1906)
40 triptorelin.ti,ab. (677)
41 arvekap.ti,ab. (1)
42 ("AY 25650" or AY25650).ti,ab. (1)
43 ("BIM 21003" or BIM21003).ti,ab. (0)
44 ("BN 52014" or BN52014).ti,ab. (0)
45 ("CL 118532" or CL118532).ti,ab. (0)
46 Debio.ti,ab. (83)
47 diphereline.ti,ab. (17)
48 moapar.ti,ab. (0)
49 pamorelin.ti,ab. (0)
50 trelstar.ti,ab. (3)
51 triptodur.ti,ab. (1)
52 ("WY 42422" or WY42422).ti,ab. (0)
53 ("WY 42462" or WY42462).ti,ab. (0)
54 gonapeptyl.ti,ab. (0)
55 decapeptyl.ti,ab. (210)
56 salvacyl.ti,ab. (0)
57 Buserelin/ (2119)
58 buserelin.ti,ab. (1304)

59 bigonist.ti,ab. (0)
60 ("hoe 766" or hoe-766 or hoe766).ti,ab. (69)
61 profact.ti,ab. (2)
62 receptal.ti,ab. (30)
63 suprecur.ti,ab. (4)
64 suprefact.ti,ab. (22)
65 tiloryth.ti,ab. (0)
66 histrelin.ti,ab. (55)
67 "LHRH-hydrogel implant".ti,ab. (1)
68 ("RL 0903" or RL0903).ti,ab. (1)
69 ("SPD 424" or SPD424).ti,ab. (1)
70 goserelin.ti,ab. (875)
71 Goserelin/ (1612)
72 ("ici 118630" or ici118630).ti,ab. (51)
73 ("ZD-9393" or ZD9393).ti,ab. (0)
74 zoladex.ti,ab. (379)
75 leuprorelin.ti,ab. (413)
76 carcinil.ti,ab. (0)
77 enanton*.ti,ab. (23)
78 ginecrin.ti,ab. (0)
79 leuplin.ti,ab. (13)
80 Leuprolide/ (2900)
81 leuprolide.ti,ab. (1743)
82 lucrin.ti,ab. (11)
83 lupron.ti,ab. (162)
84 provren.ti,ab. (0)
85 procrin.ti,ab. (3)
86 ("tap 144" or tap144).ti,ab. (40)
87 (a-43818 or a43818).ti,ab. (3)
88 Trenantone.ti,ab. (1)
89 staladex.ti,ab. (0)
90 prostap.ti,ab. (6)
91 Nafarelin/ (327)
92 nafarelin.ti,ab. (251)
93 ("76932-56-4" or "76932564").ti,ab. (0)
94 ("76932-60-0" or "76932600").ti,ab. (0)
95 ("86220-42-0" or "86220420").ti,ab. (0)
96 ("rs 94991 298" or rs94991298).ti,ab. (0)
97 synarel.ti,ab. (12)
98 deslorelin.ti,ab. (263)
99 gonadorelin.ti,ab. (201)
100 ("33515-09-2" or "33515092").ti,ab. (0)
101 ("51952-41-1" or "51952411").ti,ab. (0)
102 ("52699-48-6" or "52699486").ti,ab. (0)
103 cetrorelix.ti,ab. (463)
104 cetrotide.ti,ab. (41)
105 ("NS 75A" or NS75A).ti,ab. (0)
106 ("NS 75B" or NS75B).ti,ab. (0)

107 ("SB 075" or SB075).ti,ab. (0)
108 ("SB 75" or SB75).ti,ab. (63)
109 gonadoliberin.ti,ab. (143)
110 kryptocur.ti,ab. (6)
111 cetrorelix.ti,ab. (463)
112 cetrotide.ti,ab. (41)
113 antagon.ti,ab. (17)
114 ganirelix.ti,ab. (138)
115 ("ORG 37462" or ORG37462).ti,ab. (3)
116 orgalutran.ti,ab. (20)
117 ("RS 26306" or RS26306).ti,ab. (5)
118 ("AY 24031" or AY24031).ti,ab. (0)
119 factrel.ti,ab. (11)
120 fertagyl.ti,ab. (11)
121 lutrelef.ti,ab. (5)
122 lutrepulse.ti,ab. (3)
123 relefact.ti,ab. (10)
124 fertiral.ti,ab. (0)
125 (hoe471 or "hoe 471").ti,ab. (6)
126 relisorm.ti,ab. (4)
127 cystorelin.ti,ab. (18)
128 dirigestran.ti,ab. (5)
129 or/33-128 (42216)
130 32 and 129 (416)
131 limit 130 to english language (393)
132 limit 131 to (letter or historical article or comment or editorial or news or case reports) (36)
133 131 not 132 (357)
134 animals/ not humans/ (4686361)
135 133 not 134 (181)
136 limit 135 to yr="2000 -Current" (144)

Database: Medline in-process
Platform: Ovid
Version: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <1946 to July 21, 2020>
Search date: 23/7/2020
Number of results retrieved:
Search strategy: 42

1 Gender Dysphoria/ (0)
2 Gender Identity/ (0)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (0)
5 Transgender Persons/ (0)
6 Health Services for Transgender Persons/ (0)
7 exp Sex Reassignment Procedures/ (0)

8 (gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (1645)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (2333)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*) tw (20884)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (968)
12 (male-to-female or m2f or female-to-male or f2m).tw. (15513)
13 or/1-12 (39905)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (0)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (80723)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (0)
17 Minors/ (0)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (321871)
19 exp pediatrics/ (0)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (119783)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (0)
22 Puberty/ (0)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (60264)
24 Schools/ (0)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (0)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (69233)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (10319)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (112800)
29 or/14-28 (525529)
30 13 and 29 (9196)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (3)
32 30 or 31 (9197)
33 Gonadotropin-Releasing Hormone/ (0)
34 (pubert* adj3 block*).ti,ab. (19)
35 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (1425)
36 (GnRH adj2 analog*).ti,ab. (183)
37 GnRH*.ti,ab. (1695)
38 "GnRH agonist*".ti,ab. (379)
39 Triptorelin Pamoate/ (0)
40 triptorelin.ti,ab. (72)
41 arvekap.ti,ab. (0)
42 ("AY 25650" or AY25650).ti,ab. (0)
43 ("BIM 21003" or BIM21003).ti,ab. (0)
44 ("BN 52014" or BN52014).ti,ab. (0)
45 ("CL 118532" or CL118532).ti,ab. (0)

46 Debio.ti,ab. (11)
47 diphereline.ti,ab. (6)
48 moapar.ti,ab. (0)
49 pamorelin.ti,ab. (0)
50 trelstar.ti,ab. (0)
51 triptodur.ti,ab. (0)
52 ("WY 42422" or WY42422).ti,ab. (0)
53 ("WY 42462" or WY42462).ti,ab. (0)
54 gonapeptyl.ti,ab. (0)
55 decapeptyl.ti,ab. (8)
56 salvacyl.ti,ab. (0)
57 Buserelin/ (0)
58 buserelin.ti,ab. (59)
59 bigonist.ti,ab. (0)
60 ("hoe 766" or hoe-766 or hoe766).ti,ab. (3)
61 profact.ti,ab. (0)
62 receptal.ti,ab. (0)
63 suprecur.ti,ab. (1)
64 suprefact.ti,ab. (2)
65 tiloryth.ti,ab. (0)
66 histrelin.ti,ab. (9)
67 "LHRH-hydrogel implant".ti,ab. (0)
68 ("RL 0903" or RL0903).ti,ab. (0)
69 ("SPD 424" or SPD424).ti,ab. (0)
70 goserelin.ti,ab. (68)
71 Goserelin/ (0)
72 ("ici 118630" or ici118630).ti,ab. (0)
73 ("ZD-9393" or ZD9393).ti,ab. (0)
74 zoladex.ti,ab. (6)
75 leuprorelin.ti,ab. (47)
76 carcinil.ti,ab. (0)
77 enanton*.ti,ab. (1)
78 ginecrin.ti,ab. (0)
79 leuplin.ti,ab. (1)
80 Leuprolide/ (0)
81 leuprolide.ti,ab. (121)
82 lucrin.ti,ab. (4)
83 lupron.ti,ab. (10)
84 provren.ti,ab. (0)
85 procrin.ti,ab. (0)
86 ("tap 144" or tap144).ti,ab. (0)
87 (a-43818 or a43818).ti,ab. (0)
88 Trenantone.ti,ab. (1)
89 staladex.ti,ab. (0)
90 prostap.ti,ab. (0)
91 Nafarelin/ (0)
92 nafarelin.ti,ab. (5)
93 ("76932-56-4" or "76932564").ti,ab. (0)

94 ("76932-60-0" or "76932600").ti,ab. (0)
95 ("86220-42-0" or "86220420").ti,ab. (0)
96 ("rs 94991 298" or rs94991298).ti,ab. (0)
97 synarel.ti,ab. (0)
98 deslorelin.ti,ab. (14)
99 gonadorelin.ti,ab. (13)
100 ("33515-09-2" or "33515092").ti,ab. (0)
101 ("51952-41-1" or "51952411").ti,ab. (0)
102 ("52699-48-6" or "52699486").ti,ab. (0)
103 cetrorelix.ti,ab. (31)
104 cetrotide.ti,ab. (5)
105 ("NS 75A" or NS75A).ti,ab. (0)
106 ("NS 75B" or NS75B).ti,ab. (0)
107 ("SB 075" or SB075).ti,ab. (0)
108 ("SB 75" or SB75).ti,ab. (2)
109 gonadoliberin.ti,ab. (4)
110 kryptocur.ti,ab. (1)
111 cetrorelix.ti,ab. (31)
112 cetrotide.ti,ab. (5)
113 antagon.ti,ab. (0)
114 ganirelix.ti,ab. (8)
115 ("ORG 37462" or ORG37462).ti,ab. (0)
116 orgalutran.ti,ab. (3)
117 ("RS 26306" or RS26306).ti,ab. (0)
118 ("AY 24031" or AY24031).ti,ab. (0)
119 factrel.ti,ab. (2)
120 fertagyl.ti,ab. (1)
121 lutrelef.ti,ab. (0)
122 lutrepulse.ti,ab. (0)
123 relefact.ti,ab. (0)
124 fertiral.ti,ab. (0)
125 (hoe471 or "hoe 471").ti,ab. (0)
126 relisorm.ti,ab. (0)
127 cystorelin.ti,ab. (1)
128 dirigestran.ti,ab. (0)
129 or/33-128 (2332)
130 32 and 129 (45)
131 limit 130 to english language (45)
132 limit 131 to yr="2000 -Current" (42)

Database: Medline epubs ahead of print
Platform: Ovid
Version: Ovid MEDLINE(R) Epub Ahead of Print <July 21, 2020>
Search date: 23/7/2020
Number of results retrieved: 8
Search strategy:

1 Gender Dysphoria/ (0)

2 Gender Identity/ (0)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (0)
5 Transgender Persons/ (0)
6 Health Services for Transgender Persons/ (0)
7 exp Sex Reassignment Procedures/ (0)
8 (gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (486)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (640)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (1505)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (178)
12 (male-to-female or m2f or female-to-male or f2m).tw. (2480)
13 or/1-12 (4929)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (0)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (15496)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (0)
17 Minors/ (0)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (53563)
19 exp pediatrics/ (0)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (22796)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (0)
22 Puberty/ (0)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (13087)
24 Schools/ (0)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (0)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (12443)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (1416)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (20166)
29 or/14-28 (88366)
30 13 and 29 (1638)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (1)
32 30 or 31 (1638)
33 Gonadotropin-Releasing Hormone/ (0)
34 (pubert* adj3 block*).ti,ab. (2)
35 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (176)
36 (GnRH adj2 analog*).ti,ab. (30)
37 GnRH*.ti,ab. (223)
38 "GnRH agonist*".ti,ab. (49)
39 Triptorelin Pamoate/ (0)

40 triptorelin.ti,ab. (12)
41 arvekap.ti,ab. (0)
42 ("AY 25650" or AY25650).ti,ab. (0)
43 ("BIM 21003" or BIM21003).ti,ab. (0)
44 ("BN 52014" or BN52014).ti,ab. (0)
45 ("CL 118532" or CL118532).ti,ab. (0)
46 Debio.ti,ab. (2)
47 diphereline.ti,ab. (1)
48 moapar.ti,ab. (0)
49 pamorelin.ti,ab. (0)
50 trelstar.ti,ab. (0)
51 triptodur.ti,ab. (0)
52 ("WY 42422" or WY42422).ti,ab. (0)
53 ("WY 42462" or WY42462).ti,ab. (0)
54 gonapeptyl.ti,ab. (0)
55 decapeptyl.ti,ab. (0)
56 salvacyl.ti,ab. (0)
57 Buserelin/ (0)
58 buserelin.ti,ab. (7)
59 bigonist.ti,ab. (0)
60 ("hoe 766" or hoe-766 or hoe766).ti,ab. (0)
61 profact.ti,ab. (0)
62 receptal.ti,ab. (0)
63 suprecur.ti,ab. (0)
64 suprefact.ti,ab. (1)
65 tiloryth.ti,ab. (0)
66 histrelin.ti,ab. (2)
67 "LHRH-hydrogel implant".ti,ab. (0)
68 ("RL 0903" or RL0903).ti,ab. (0)
69 ("SPD 424" or SPD424).ti,ab. (0)
70 goserelin.ti,ab. (11)
71 Goserelin/ (0)
72 ("ici 118630" or ici118630).ti,ab. (0)
73 ("ZD-9393" or ZD9393).ti,ab. (0)
74 zoladex.ti,ab. (1)
75 leuprorelin.ti,ab. (13)
76 carcinil.ti,ab. (0)
77 enanton*.ti,ab. (1)
78 ginecrin.ti,ab. (0)
79 leuplin.ti,ab. (0)
80 Leuprolide/ (0)
81 leuprolide.ti,ab. (22)
82 lucrin.ti,ab. (0)
83 lupron.ti,ab. (2)
84 provren.ti,ab. (0)
85 procrin.ti,ab. (0)
86 ("tap 144" or tap144).ti,ab. (1)
87 (a-43818 or a43818).ti,ab. (0)

88 Trenantone.ti,ab. (0)
89 staladex.ti,ab. (0)
90 prostap.ti,ab. (0)
91 Nafarelin/ (0)
92 nafarelin.ti,ab. (4)
93 ("76932-56-4" or "76932564").ti,ab. (0)
94 ("76932-60-0" or "76932600").ti,ab. (0)
95 ("86220-42-0" or "86220420").ti,ab. (0)
96 ("rs 94991 298" or rs94991298).ti,ab. (0)
97 synarel.ti,ab. (0)
98 deslorelin.ti,ab. (3)
99 gonadorelin.ti,ab. (3)
100 ("33515-09-2" or "33515092").ti,ab. (0)
101 ("51952-41-1" or "51952411").ti,ab. (0)
102 ("52699-48-6" or "52699486").ti,ab. (0)
103 cetrorelix.ti,ab. (6)
104 cetrotide.ti,ab. (2)
105 ("NS 75A" or NS75A).ti,ab. (0)
106 ("NS 75B" or NS75B).ti,ab. (0)
107 ("SB 075" or SB075).ti,ab. (0)
108 ("SB 75" or SB75).ti,ab. (0)
109 gonadoliberin.ti,ab. (0)
110 kryptocur.ti,ab. (0)
111 cetrorelix.ti,ab. (6)
112 cetrotide.ti,ab. (2)
113 antagon.ti,ab. (1)
114 ganirelix.ti,ab. (1)
115 ("ORG 37462" or ORG37462).ti,ab. (0)
116 orgalutran.ti,ab. (0)
117 ("RS 26306" or RS26306).ti,ab. (0)
118 ("AY 24031" or AY24031).ti,ab. (0)
119 factrel.ti,ab. (0)
120 fertagyl.ti,ab. (0)
121 lutrelef.ti,ab. (0)
122 lutrepulse.ti,ab. (0)
123 relefact.ti,ab. (0)
124 fertiral.ti,ab. (0)
125 (hoe471 or "hoe 471").ti,ab. (0)
126 relisorm.ti,ab. (0)
127 cystorelin.ti,ab. (0)
128 dirigestran.ti,ab. (0)
129 or/33-128 (310)
130 32 and 129 (8)
131 limit 130 to english language (8)
132 limit 131 to yr="2000 -Current" (8)

Database: Medline daily update
Platform: Ovid

Version: Ovid MEDLINE(R) Daily Update <July 21, 2020>
Search date: 23/7/2020
Number of results retrieved: 1
Search strategy

1 Gender Dysphoria/ (4)
2 Gender Identity/ (38)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (2)
5 Transgender Persons/ (26)
6 Health Services for Transgender Persons/ (1)
7 exp Sex Reassignment Procedures/ (3)
8 (gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (24)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (39)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (87)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (15)
12 (male-to-female or m2f or female-to-male or f2m).tw. (181)
13 or/1-12 (358)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (932)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (981)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (1756)
17 Minors/ (3)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (3672)
19 exp pediatrics/ (75)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (1658)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (2006)
22 Puberty/ (8)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (732)
24 Schools/ (56)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (5)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (622)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (98)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (1301)
29 or/14-28 (6705)
30 13 and 29 (130)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (0)
32 30 or 31 (130)
33 Gonadotropin-Releasing Hormone/ (11)

34 (pubert* adj3 block*).ti,ab. (0)
35 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (10)
36 (GnRH adj2 analog*).ti,ab. (2)
37 GnRH*.ti,ab. (14)
38 "GnRH agonist*".ti,ab. (4)
39 Triptorelin Pamoate/ (1)
40 triptorelin.ti,ab. (1)
41 arvekap.ti,ab. (0)
42 ("AY 25650" or AY25650).ti,ab. (0)
43 ("BIM 21003" or BIM21003).ti,ab. (0)
44 ("BN 52014" or BN52014).ti,ab. (0)
45 ("CL 118532" or CL118532).ti,ab. (0)
46 Debio.ti,ab. (1)
47 diphereline.ti,ab. (0)
48 moapar.ti,ab. (0)
49 pamorelin.ti,ab. (0)
50 trelstar.ti,ab. (0)
51 triptodur.ti,ab. (0)
52 ("WY 42422" or WY42422).ti,ab. (0)
53 ("WY 42462" or WY42462).ti,ab. (0)
54 gonapeptyl.ti,ab. (0)
55 decapeptyl.ti,ab. (0)
56 salvacyl.ti,ab. (0)
57 Buserelin/ (0)
58 buserelin.ti,ab. (0)
59 bigonist.ti,ab. (0)
60 ("hoe 766" or hoe-766 or hoe766).ti,ab. (0)
61 profact.ti,ab. (0)
62 receptal.ti,ab. (0)
63 suprecur.ti,ab. (0)
64 suprefact.ti,ab. (0)
65 tiloryth.ti,ab. (0)
66 histrelin.ti,ab. (0)
67 "LHRH-hydrogel implant".ti,ab. (0)
68 ("RL 0903" or RL0903).ti,ab. (0)
69 ("SPD 424" or SPD424).ti,ab. (0)
70 goserelin.ti,ab. (1)
71 Goserelin/ (2)
72 ("ici 118630" or ici118630).ti,ab. (0)
73 ("ZD-9393" or ZD9393).ti,ab. (0)
74 zoladex.ti,ab. (0)
75 leuprorelin.ti,ab. (0)
76 carcinil.ti,ab. (0)
77 enanton*.ti,ab. (0)
78 ginecrin.ti,ab. (0)
79 leuplin.ti,ab. (0)
80 Leuprolide/ (0)
81 leuprolide.ti,ab. (0)

82 lucrin.ti,ab. (0)
83 lupron.ti,ab. (0)
84 provren.ti,ab. (0)
85 procrin.ti,ab. (0)
86 ("tap 144" or tap144).ti,ab. (0)
87 (a-43818 or a43818).ti,ab. (0)
88 Trenantone.ti,ab. (0)
89 staladex.ti,ab. (0)
90 prostap.ti,ab. (0)
91 Nafarelin/ (0)
92 nafarelin.ti,ab. (0)
93 ("76932-56-4" or "76932564").ti,ab. (0)
94 ("76932-60-0" or "76932600").ti,ab. (0)
95 ("86220-42-0" or "86220420").ti,ab. (0)
96 ("rs 94991 298" or rs94991298).ti,ab. (0)
97 synarel.ti,ab. (0)
98 deslorelin.ti,ab. (0)
99 gonadorelin.ti,ab. (0)
100 ("33515-09-2" or "33515092").ti,ab. (0)
101 ("51952-41-1" or "51952411").ti,ab. (0)
102 ("52699-48-6" or "52699486").ti,ab. (0)
103 cetrorelix.ti,ab. (0)
104 cetrotide.ti,ab. (0)
105 ("NS 75A" or NS75A).ti,ab. (0)
106 ("NS 75B" or NS75B).ti,ab. (0)
107 ("SB 075" or SB075).ti,ab. (0)
108 ("SB 75" or SB75).ti,ab. (0)
109 gonadoliberin.ti,ab. (0)
110 kryptocur.ti,ab. (0)
111 cetrorelix.ti,ab. (0)
112 cetrotide.ti,ab. (0)
113 antagon.ti,ab. (0)
114 ganirelix.ti,ab. (0)
115 ("ORG 37462" or ORG37462).ti,ab. (0)
116 orgalutran.ti,ab. (0)
117 ("RS 26306" or RS26306).ti,ab. (0)
118 ("AY 24031" or AY24031).ti,ab. (0)
119 factrel.ti,ab. (0)
120 fertagyl.ti,ab. (0)
121 lutrelef.ti,ab. (0)
122 lutrepulse.ti,ab. (0)
123 relefact.ti,ab. (0)
124 fertiral.ti,ab. (0)
125 (hoe471 or "hoe 471").ti,ab. (0)
126 relisorm.ti,ab. (0)
127 cystorelin.ti,ab. (0)
128 dirigestran.ti,ab. (0)
129 or/33-128 (23)

130 32 and 129 (1)
131 limit 130 to english language (1)
132 limit 131 to yr="2000 -Current" (1)

Database: Embase
Platform: Ovid
Version: Embase <1974 to 2020 July 22>
Search date: 23/7/2020
Number of results retrieved: 367
Search strategy:

1 exp Gender Dysphoria/ (5399)
2 Gender Identity/ (16820)
3 "Sexual and Gender Disorders"/ (24689)
4 Transsexualism/ (3869)
5 exp Transgender/ (6597)
6 Health Services for Transgender Persons/ (158848)
7 exp Sex Reassignment Procedures/ or sex transformation/ (3058)
8 (gender* adj3 (dysphori* or affirm* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (13005)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (22509)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (154446)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (10327)
12 (male-to-female or m2f or female-to-male or f2m).tw. (200166)
13 or/1-12 (582812)
14 exp juvenile/ or Child Behavior/ or Child Welfare/ or Child Health/ or infant welfare/ or "minor (person)"/ or elementary student/ (3437324)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (1186161)
16 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (3586795)
17 exp pediatrics/ (106214)
18 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (1491597)
19 exp adolescence/ or exp adolescent behavior/ or adolescent health/ or high school student/ or middle school student/ (105108)
20 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (641660)
21 school/ or high school/ or kindergarten/ or middle school/ or primary school/ or nursery school/ or day care/ (103791)
22 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (687437)
23 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (138908)
24 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (1562903)

25 or/14-24 (7130881)
26 13 and 25 (182161)
27 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw.
(17)
28 26 or 27 (182161)
29 gonadorelin/ (37580)
30 (pubert* adj3 block*).ti,ab. (142)
31 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (21450)
32 (GnRH adj2 analog*).ti,ab. (4013)
33 GnRH*.ti,ab. (29862)
34 "GnRH agonist*".ti,ab. (6719)
35 exp gonadorelin agonist/ or gonadorelin derivative/ or gonadorelin acetate/ (23304)
36 Triptorelin/ (5427)
37 triptorelin.ti,ab. (1182)
38 arvekap.ti,ab. (3)
39 ("AY 25650" or AY25650).ti,ab. (1)
40 ("BIM 21003" or BIM21003).ti,ab. (0)
41 ("BN 52014" or BN52014).ti,ab. (0)
42 ("CL 118532" or CL118532).ti,ab. (0)
43 Debio.ti,ab. (185)
44 diphereline.ti,ab. (51)
45 moapar.ti,ab. (0)
46 pamorelin.ti,ab. (0)
47 trelstar.ti,ab. (5)
48 triptodur.ti,ab. (1)
49 ("WY 42422" or WY42422).ti,ab. (0)
50 ("WY 42462" or WY42462).ti,ab. (0)
51 gonapeptyl.ti,ab. (10)
52 decapeptyl.ti,ab. (307)
53 salvacyl.ti,ab. (1)
54 buserelin acetate/ or buserelin/ (5164)
55 buserelin.ti,ab. (1604)
56 bigonist.ti,ab. (1)
57 ("hoe 766" or hoe-766 or hoe766).ti,ab. (89)
58 profact.ti,ab. (4)
59 receptal.ti,ab. (37)
60 suprecur.ti,ab. (8)
61 suprefact.ti,ab. (30)
62 tiloryth.ti,ab. (0)
63 histrelin/ (446)
64 histrelin.ti,ab. (107)
65 "LHRH-hydrogel implant".ti,ab. (1)
66 ("RL 0903" or RL0903).ti,ab. (1)
67 ("SPD 424" or SPD424).ti,ab. (1)
68 goserelin.ti,ab. (1487)
69 Goserelin/ (7128)
70 ("ici 118630" or ici118630).ti,ab. (49)
71 ("ZD-9393" or ZD9393).ti,ab. (0)

72 zoladex.ti,ab. (501)
73 leuprorelin/ (11312)
74 leuprorelin.ti,ab. (727)
75 carcinil.ti,ab. (0)
76 enanton*.ti,ab. (38)
77 ginecrin.ti,ab. (1)
78 leuplin.ti,ab. (26)
79 leuprolide.ti,ab. (2788)
80 lucrin.ti,ab. (47)
81 lupron.ti,ab. (361)
82 provren.ti,ab. (0)
83 procrin.ti,ab. (11)
84 ("tap 144" or tap144).ti,ab. (63)
85 (a-43818 or a43818).ti,ab. (3)
86 Trenantone.ti,ab. (7)
87 staladex.ti,ab. (0)
88 prostap.ti,ab. (11)
89 nafarelin acetate/ or nafarelin/ (1441)
90 nafarelin.ti,ab. (324)
91 ("76932-56-4" or "76932564").ti,ab. (0)
92 ("76932-60-0" or "76932600").ti,ab. (0)
93 ("86220-42-0" or "86220420").ti,ab. (0)
94 ("rs 94991 298" or rs94991298).ti,ab. (0)
95 synarel.ti,ab. (28)
96 deslorelin/ (452)
97 deslorelin.ti,ab. (324)
98 gonadorelin.ti,ab. (338)
99 ("33515-09-2" or "33515092").ti,ab. (0)
100 ("51952-41-1" or "51952411").ti,ab. (0)
101 ("52699-48-6" or "52699486").ti,ab. (0)
102 cetrorelix/ (2278)
103 cetrorelix.ti,ab. (717)
104 cetrotide.ti,ab. (113)
105 ("NS 75A" or NS75A).ti,ab. (0)
106 ("NS 75B" or NS75B).ti,ab. (0)
107 ("SB 075" or SB075).ti,ab. (1)
108 ("SB 75" or SB75).ti,ab. (76)
109 gonadoliberin.ti,ab. (152)
110 kryptocur.ti,ab. (6)
111 cetrorelix.ti,ab. (717)
112 cetrotide.ti,ab. (113)
113 antagon.ti,ab. (32)
114 ganirelix/ (1284)
115 ganirelix.ti,ab. (293)
116 ("ORG 37462" or ORG37462).ti,ab. (4)
117 orgalutran/ (1284)
118 orgalutran.ti,ab. (68)
119 ("RS 26306" or RS26306).ti,ab. (6)

120 ("AY 24031" or AY24031).ti,ab. (0)
121 factrel.ti,ab. (14)
122 fertagyl.ti,ab. (20)
123 lutrelef.ti,ab. (7)
124 lutrepulse.ti,ab. (6)
125 relefact.ti,ab. (10)
126 fertiral.ti,ab. (0)
127 (hoe471 or "hoe 471").ti,ab. (4)
128 relisorm.ti,ab. (6)
129 cystorelin.ti,ab. (26)
130 dirigestran.ti,ab. (5)
131 or/29-130 (80790)
132 28 and 131 (988)
133 limit 132 to english language (940)
134 133 not (letter or editorial).pt. (924)
135 134 not (conference abstract or conference paper or conference proceeding or "conference review").pt. (683)
136 nonhuman/ not (human/ and nonhuman/) (4649157)
137 135 not 136 (506)
138 limit 137 to yr="2000 -Current" (420)
139 elsevier.cr. (25912990)
140 138 and 139 (372)
141 remove duplicates from 140 (367)

Database: Cochrane Library – incorporating Cochrane Database of Systematic Reviews (CDSR); CENTRAL
Platform: Wiley
Version:
CDSR – Issue 7 of 12, July 2020
CENTRAL – Issue 7 of 12, July 2020
Search date: 23/7/2020
Number of results retrieved: CDSR – 1; CENTRAL - 8.

#1 [mh ^"Gender Dysphoria"] 3
#2 [mh ^"gender identity"] 227
#3 [mh ^"sexual and gender disorders"] 2
#4 [mh ^transsexualism] 27
#5 [mh ^"transgender persons"] 36
#6 [mh ^"health services for transgender persons"] 0
#7 [mh "sex reassignment procedures"] 4
#8 (gender* NEAR/3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)):ti,ab 308
#9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*):ti,ab 929
#10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*):ti,ab 3915
#11 ((sex or gender*) NEAR/3 (reassign* or chang* or transform* or transition*)):ti,ab 493
#12 (male-to-female or m2f or female-to-male or f2m):ti,ab 489

#13 {or #1-#12} 6142
#14 [mh infant] or [mh ^"infant health"] or [mh ^"infant welfare"] 27769
#15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*):ti,ab 69476
#16 [mh child] or [mh "child behavior"] or [mh ^"child health"] or [mh ^"child welfare"] 42703
#17 [mh ^minors] 8
#18 (child* or minor or minors or boy* or girl* or kid or kids or young*):ti,ab 175826
#19 [mh pediatrics]661
#20 (pediatric* or paediatric* or peadiatric*):ti,ab 30663
#21 [mh ^adolescent] or [mh ^"adolescent behavior"] or [mh ^"adolescent health"] 102154
#22 [mh ^puberty] 295
#23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*):ti,ab 34139
#24 [mh ^schools] 1914
#25 [mh ^"Child Day Care Centers"] or [mh nurseries] or [mh ^"schools, nursery"] 277
#26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*):ti,ab 54723
#27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") NEAR/2 (year or years or age or ages or aged)):ti,ab 6710
#28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") NEAR/2 (year or years or age or ages or aged)):ti,ab 196881
#29 {or #14-#28} 469351
#30 #13 and #29 2146
#31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*):ti,ab 0
#32 #30 or #31 2146
#33 [mh ^"Gonadotropin-Releasing Hormone"] 1311
#34 (pubert* NEAR/3 block*):ti,ab 1
#35 ((gonadotrophin or gonadotropin) and releasing):ti,ab 2095
#36 (GnRH NEAR/2 analog*):ti,ab 493
#37 GnRH*:ti,ab 3764
#38 "GnRH agonist*":ti,ab 1399
#39 [mh ^"Triptorelin Pamoate"] 451
#40 triptorelin:ti,ab 451
#41 arvekap:ti,ab 4
#42 ("AY 25650" or AY25650):ti,ab 0
#43 ("BIM 21003" or BIM21003):ti,ab 0
#44 ("BN 52014" or BN52014):ti,ab 0
#45 ("CL 118532" or CL118532):ti,ab 0
#46 Debio:ti,ab 301
#47 diphereline:ti,ab 25
#48 moapar:ti,ab 0
#49 pamorelin:ti,ab 5
#50 trelstar:ti,ab 3

#51 triptodur:ti,ab 0
#52 ("WY 42422" or WY42422):ti,ab 0
#53 ("WY 42462" or WY42462):ti,ab 0
#54 gonapeptyl:ti,ab 11
#55 decapeptyl:ti,ab 135
#56 salvacyl:ti,ab 0
#57 [mh ^Buserelin] 290
#58 Buserelin:ti,ab 339
#59 bigonist:ti,ab 0
#60 ("hoe 766" or hoe-766 or hoe766):ti,ab 11
#61 profact:ti,ab 1
#62 receptal:ti,ab 4
#63 suprecur:ti,ab 0
#64 suprefact:ti,ab 28
#65 tiloryth:ti,ab 0
#66 histrelin:ti,ab 5
#67 "LHRH-hydrogel implant":ti,ab 0
#68 ("RL 0903" or RL0903):ti,ab 0
#69 ("SPD 424" or SPD424):ti,ab 0
#70 goserelin:ti,ab 761
#71 [mh ^goserelin] 568
#72 ("ici 118630" or ici118630):ti,ab 7
#73 ("ZD-9393" or ZD9393):ti,ab 1
#74 zoladex:ti,ab 318
#75 leuprorelin:ti,ab 248
#76 carcinil:ti,ab 0
#77 enanton*:ti,ab 21
#78 ginecrin:ti,ab 1
#79 leuplin:ti,ab 7
#80 [mh ^Leuprolide] 686
#81 leuprolide:ti,ab696
#82 lucrin:ti,ab 21
#83 lupron:ti,ab 77
#84 provren:ti,ab 0
#85 procrin:ti,ab 2
#86 ("tap 144" or tap144):ti,ab 24
#87 (a-43818 or a43818):ti,ab 0
#88 Trenantone:ti,ab 3
#89 staladex:ti,ab 0
#90 prostap:ti,ab 9
#91 [mh ^Nafarelin] 77
#92 nafarelin:ti,ab 114
#93 ("76932-56-4" or "76932564"):ti,ab 0
#94 ("76932-60-0" or "76932600"):ti,ab 2
#95 ("86220-42-0" or "86220420"):ti,ab 0
#96 ("rs 94991 298" or rs94991298):ti,ab 0
#97 synarel:ti,ab 10
#98 deslorelin:ti,ab16

#99 gonadorelin:ti,ab 11
#100 ("33515-09-2" or "33515092"):ti,ab 0
#101 ("51952-41-1" or "51952411"):ti,ab 0
#102 ("52699-48-6" or "52699486"):ti,ab 0
#103 cetrorelix:ti,ab 221
#104 cetrotide:ti,ab 111
#105 ("NS 75A" or NS75A):ti,ab 0
#106 ("NS 75B" or NS75B):ti,ab 0
#107 ("SB 075" or SB075):ti,ab 0
#108 ("SB 75" or SB75):ti,ab 10
#109 gonadoliberin:ti,ab 5
#110 kryptocur:ti,ab 0
#111 cetrorelix:ti,ab 221
#112 cetrotide:ti,ab 111
#113 antagon:ti,ab 12
#114 ganirelix:ti,ab 142
#115 ("ORG 37462" or ORG37462):ti,ab 4
#116 orgalutran:ti,ab 45
#117 ("RS 26306" or RS26306):ti,ab 0
#118 ("AY 24031" or AY24031):ti,ab 0
#119 factrel:ti,ab 1
#120 fertagyl:ti,ab 0
#121 lutrelef:ti,ab 0
#122 lutrepulse:ti,ab 1
#123 relefact:ti,ab 1
#124 fertiral:ti,ab 0
#125 (hoe471 or "hoe 471"):ti,ab 3
#126 relisorm:ti,ab 0
#127 cystorelin:ti,ab 0
#128 dirigestran:ti,ab 0
#129 {or #33-#128} 6844
#130 #32 and #129 27
#131 #130 with Cochrane Library publication date Between Jan 2000 and Jul 2020, in Cochrane Reviews 1
#132 #130 27
#133 "conference":pt or (clinicaltrials or trialsearch):so 492465
#134 #132 not #133 9
#135 #134 with Publication Year from 2000 to 2020, in Trials 8

Database: HTA
Platform: CRD
Version: HTA
Search date: 23/7/2020
Number of results retrieved: 26
Search strategy:

1 MeSH DESCRIPTOR Gender Dysphoria EXPLODE ALL TREES 0
2 MeSH DESCRIPTOR Gender Identity EXPLODE ALL TREES 14

3 MeSH DESCRIPTOR Sexual and Gender Disorders EXPLODE ALL TREES 2

4 MeSH DESCRIPTOR Transsexualism EXPLODE ALL TREES 12
5 MeSH DESCRIPTOR Transgender Persons EXPLODE ALL TREES 3
6 MeSH DESCRIPTOR Health Services for Transgender Persons EXPLODE ALL TREES 0
7 MeSH DESCRIPTOR Sex Reassignment Procedures EXPLODE ALL TREES 1

8 ((gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*))) 28
9 ((transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*)) 76
10 ((trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*)) 83
11 (((sex or gender*) adj3 (reassign* or chang* or transform* or transition*))) 24
12 (male-to-female or m2f or female-to-male or f2m) 86
13 ((transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*)) 0
14 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 262
15 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13) IN HTA 30

*26 results are from 200 onwards. Downloaded as a set to sift for drug terms rather than continuing with search strategy.

Database: APA PsycInfo
Search date: July 2020 (Week 2)
Search Strategy:

---

1 Gender Dysphoria/ (936)
2 Gender Identity/ (8648)
3 Transsexualism/ (2825)
4 Transgender/ (5257)
5 exp Gender Reassignment/ (568)
6 (gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (15471)
7 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (13028)
8 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (7679)
9 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (5796)
10 (male-to-female or m2f or female-to-male or f2m).tw. (63688)
11 or/1-10 (99560)
12 exp Infant Development/ (21841)
13 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (150219)

14 Child Characteristics/ or exp Child Behavior/ or Child Psychology/ or exp Child Welfare/ or Child Psychiatry/ (23423)
15 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (984230)
16 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (78962)
17 Adolescent Psychiatry/ or Adolescent Behavior/ or Adolescent Development/ or Adolescent Psychology/ or Adolescent Characteristics/ or Adolescent Health/ (62142)
18 Puberty/ (2753)
19 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (347604)
20 Schools/ or exp elementary school students/ or high school students/ or junior high school students/ or middle school students/ (113053)
21 Child Day Care/ or Nursery Schools/ (2836)
22 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (772814)
23 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (21475)
24 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (285697)
25 or/12-24 (1772959)
26 11 and 25 (49612)
27 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (14)
28 26 or 27 (49613)
29 exp Gonadotropic Hormones/ (4226)
30 (pubert* adj3 block*).ti,ab. (29)
31 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (1060)
32 (GnRH adj2 analog*).ti,ab. (49)
33 GnRH*.ti,ab. (998)
34 "GnRH agonist*".ti,ab. (72)
35 triptorelin.ti,ab. (25)
36 arvekap.ti,ab. (0)
37 ("AY 25650" or AY25650).ti,ab. (0)
38 ("BIM 21003" or BIM21003).ti,ab. (0)
39 ("BN 52014" or BN52014).ti,ab. (0)
40 ("CL 118532" or CL118532).ti,ab. (0)
41 Debio.ti,ab. (7)
42 diphereline.ti,ab. (0)
43 moapar.ti,ab. (0)
44 pamorelin.ti,ab. (0)
45 trelstar.ti,ab. (0)
46 triptodur.ti,ab. (0)
47 ("WY 42422" or WY42422).ti,ab. (0)
48 ("WY 42462" or WY42462).ti,ab. (0)
49 gonapeptyl.ti,ab. (0)
50 decapeptyl.ti,ab. (3)
51 salvacyl.ti,ab. (1)

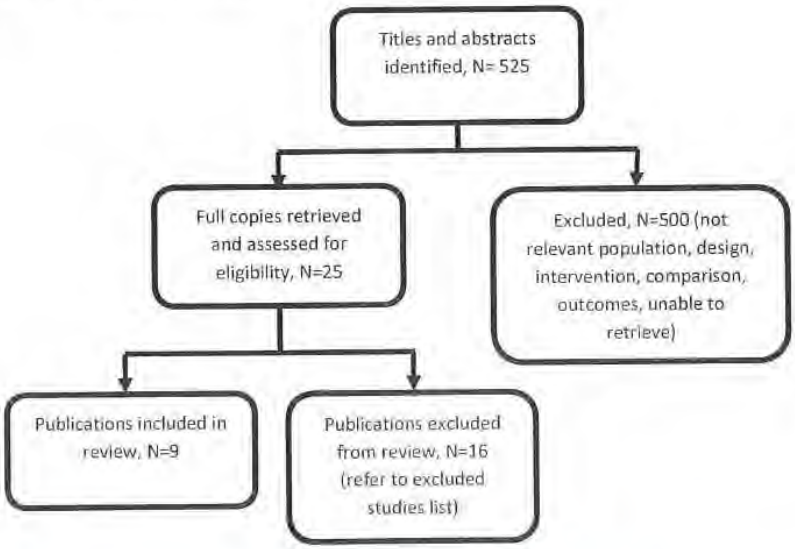
52 buserelin.ti,ab. (6)
53 bigonist.ti,ab. (0)
54 ("hoe 766" or hoe-766 or hoe766).ti,ab. (0)
55 profact.ti,ab. (0)
56 receptal.ti,ab. (0)
57 suprecur.ti,ab. (0)
58 suprefact.ti,ab. (0)
59 tiloryth.ti,ab. (0)
60 histrelin.ti,ab. (1)
61 "LHRH-hydrogel implant".ti,ab. (0)
62 ("RL 0903" or RL0903).ti,ab. (0)
63 ("SPD 424" or SPD424).ti,ab. (0)
64 goserelin.ti,ab. (30)
65 ("ici 118630" or ici118630).ti,ab. (0)
66 ("ZD-9393" or ZD9393).ti,ab. (0)
67 zoladex.ti,ab. (3)
68 leuprorelin.ti,ab. (12)
69 carcinil.ti,ab. (0)
70 enanton*.ti,ab. (1)
71 ginecrin.ti,ab. (0)
72 leuplin.ti,ab. (0)
73 leuprolide.ti,ab. (79)
74 lucrin.ti,ab. (1)
75 lupron.ti,ab. (18)
76 provren.ti,ab. (0)
77 procrin.ti,ab. (0)
78 ("tap 144" or tap144).ti,ab. (1)
79 (a-43818 or a43818).ti,ab. (0)
80 Trenantone.ti,ab. (0)
81 staladex.ti,ab. (0)
82 prostap.ti,ab. (0)
83 nafarelin.ti,ab. (1)
84 ("76932-56-4" or "76932564").ti,ab. (0)
85 ("76932-60-0" or "76932600").ti,ab. (0)
86 ("86220-42-0" or "86220420").ti,ab. (0)
87 ("rs 94991 298" or rs94991298).ti,ab. (0)
88 synarel.ti,ab. (0)
89 deslorelin.ti,ab. (8)
90 gonadorelin.ti,ab. (3)
91 ("33515-09-2" or "33515092").ti,ab. (0)
92 ("51952-41-1" or "51952411").ti,ab. (0)
93 ("52699-48-6" or "52699486").ti,ab. (0)
94 cetrorelix.ti,ab. (9)
95 cetrotide.ti,ab. (0)
96 ("NS 75A" or NS75A).ti,ab. (0)
97 ("NS 75B" or NS75B).ti,ab. (0)
98 ("SB 075" or SB075).ti,ab. (0)
99 ("SB 75" or SB75).ti,ab. (1)

100	gonadoliberin.ti,ab. (1)
101	kryptocur.ti,ab. (0)
102	cetrorelix.ti,ab. (9)
103	cetrotide.ti,ab. (0)
104	antagon.ti,ab. (0)
105	ganirelix.ti,ab. (0)
106	("ORG 37462" or ORG37462).ti,ab. (0)
107	orgalutran.ti,ab. (0)
108	("RS 26306" or RS26306).ti,ab. (0)
109	("AY 24031" or AY24031).ti,ab. (0)
110	factrel.ti,ab. (0)
111	fertagyl.ti,ab. (0)
112	lutrelef.ti,ab. (0)
113	lutrepulse.ti,ab. (0)
114	relefact.ti,ab. (0)
115	fertiral.ti,ab. (0)
116	(hoe471 or "hoe 471").ti,ab. (0)
117	relisorm.ti,ab. (0)
118	cystorelin.ti,ab. (0)
119	dirigestran.ti,ab. (0)
120	or/29-119 (4869)
121	28 and 120 (130)
122	limit 121 to english language (120)
123	limit 122 to yr="2000 -Current" (93)

Appendix C Evidence selection

The literature searches identified 525 references. These were screened using their titles and abstracts and 25 references were obtained and assessed for relevance. Of these, 9 references are included in the evidence review. The remaining 16 references were excluded and are listed in appendix D.

Figure 1 – Study selection flow diagram



References submitted with Preliminary Policy Proposal

There is no preliminary policy proposal for this policy.

Appendix D Excluded studies table

Study reference	Reason for exclusion
Achille, C., Taggart, T., Eaton, N.R. et al. (2020) Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: Preliminary results. International Journal of Pediatric Endocrinology 2020(1): 8	Intervention – data for GnRH analogues not reported separately from other interventions
Bechard, Melanie, Vanderlaan, Doug P, Wood, Hayley et al. (2017) Psychosocial and Psychological Vulnerability in Adolescents with Gender Dysphoria: A "Proof of Principle" Study. Journal of sex & marital therapy 43(7): 678-688	Population – no GnRH analogues at time of study
Chew, Denise, Anderson, Jemma, Williams, Katrina et al. (2018) Hormonal Treatment in Young People With Gender Dysphoria: A Systematic Review. Pediatrics 141(4)	All primary studies included apart from 1 conference abstract
de Vries, Annelou L C, McGuire, Jenifer K et al. (2014) Young adult psychological outcome after puberty suppression and gender reassignment. Pediatrics 134(4): 696-704	Population – relevant population included in de Vries et al. 2011
Ghelani, Rahul, Lim, Cheryl, Brain, Caroline et al. (2020) Sudden sex hormone withdrawal and the effects on body composition in late pubertal adolescents with gender dysphoria. Journal of pediatric endocrinology & metabolism: JPEM 33(1): 107-112	Outcomes – not in the PICO

Study reference	Reason for exclusion
Giovanardi, G, Morales, P, Mirabella, M et al. (2019) Transition memories: experiences of trans adult women with hormone therapy and their beliefs on the usage of hormone blockers to suppress puberty. Journal of endocrinological investigation 42(10): 1231-1240	Population – adults only
Hewitt, Jacqueline K, Paul, Campbell, Kasiannan, Porpavai et al. (2012) Hormone treatment of gender identity disorder in a cohort of children and adolescents. The Medical journal of Australia 196(9): 578-81	Outcomes – no data reported for relevant outcomes
Jensen, R.K., Jensen, J.K., Simons, L.K. et al. (2019) Effect of Concurrent Gonadotropin-Releasing Hormone Agonist Treatment on Dose and Side Effects of Gender-Affirming Hormone Therapy in Adolescent Transgender Patients. Transgender Health 4(1): 300-303	Outcomes – not in the PICO
Klaver, Maartje, de Mutsert, Renee, Wiepjes, Chantal M et al. (2018) Early Hormonal Treatment Affects Body Composition and Body Shape in Young Transgender Adolescents. The journal of sexual medicine 15(2): 251-260	Outcomes – not in the PICO
Klaver, Maartje, de Mutsert, Renee van der Loos, Maria A T C et al. (2020) Hormonal Treatment and Cardiovascular Risk Profile in Transgender Adolescents. Pediatrics 145(3)	Outcomes – not in the PICO
Lopez, Carla Marisa, Solomon, Daniel, Boulware, Susan D et al. (2018) Trends in the use of puberty blockers among transgender children in the United States. Journal of pediatric endocrinology & metabolism : JPEM 31(6): 665-670	Outcomes – not in the PICO
Schagen, Sebastian E E, Lustenhouwer, Paul, Cohen-Kettenis, Peggy T et al. (2018) Changes in Adrenal Androgens During Puberty Suppression and Gender-Affirming Hormone Treatment in Adolescents With Gender Dysphoria. The journal of sexual medicine 15(9): 1357-1363	Outcomes – not in the PICO
Swendiman, Robert A, Vogiatzi, Maria G, Alter, Craig A et al. (2019) Histrelin implantation in the pediatric population: A 10-year institutional experience. Journal of pediatric surgery 54(7): 1457-1461	Population – less than 10% of participants had gender dysphoria, data not reported separately
Turban, Jack L, King, Dana, Carswell, Jeremi M et al. (2020) Pubertal Suppression for Transgender Youth and Risk of Suicidal Ideation. Pediatrics 145(2)	Intervention – data for GnRH analogues not reported separately from other interventions
Vrouenraets, Lieke Josephina Jeanne Johanna, Fredriks, A Miranda, Hannema, Sabine E et al. (2016) Perceptions of Sex, Gender, and Puberty Suppression: A Qualitative Analysis of Transgender Youth. Archives of sexual behavior 45(7): 1697-703	Outcomes – not in the PICO
Zucker, Kenneth J, Bradley, Susan J, Owen-Anderson, Allison et al. (2010) Puberty-blocking hormonal therapy for adolescents with gender identity disorder: A descriptive clinical study. Journal of Gay & Lesbian Mental Health 15(1): 58-82	Intervention – data for GnRH analogues not reported separately from other interventions

Appendix E Evidence tables

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Brik T, Vrouenraets L, de Vries M, et al. (2020) Trajectories of adolescents treated with gonadotropin-releasing hormone analogues for gender dysphoria. Archives of Sexual Behaviour https://doi.org/10.1007/s10508-020-01660-8 Netherlands Retrospective observational single-centre study To document trajectories after the initiation of GnRH analogue and explore reasons for extended use and discontinuation of GnRH analogues. Includes participants seen between November 2010 and January 1, 2018.	Inclusion criteria were adolescents with gender dysphoria, according to the DSM-5 criteria, seen at the single centre and treated with GnRH analogues between November 2010 and January 1, 2018. The study excluded adolescents without a diagnosis of gender dysphoria, those who had coexisting problems that interfered with the diagnostic process and/or might interfere with successful treatment (not further defined), those adolescents not wanting hormones, those with ongoing diagnostic evaluation and those who did not attend appointments. The sample consisted of 143 adolescents meeting the inclusion/exclusion criteria, 38 transfemales, 105 transmales, with median ages of 15.0 years (range 11.1 to 18.6 years) and 16.1 years (range 10.1 to 17.9	The study only reports that GnRH analogues were given, no specific drug, dose, route, or frequency of administration are reported. No comparator cohort was used in the study. Follow-up was at (up to) 9 years (last follow-up July 2019).	**Critical outcomes** No critical outcomes assessed. **Important outcomes** ***Psychosocial impact*** Not assessed. ***Engagement with health care services*** Not formally assessed but the study reported that out of 214 age and developmentally appropriate adolescents for potential inclusion in the study, 9 were excluded as they stopped attending appointments (4.2%). ***Stopping treatment*** Of the 143 adolescents, 9 (6.2%, 1 transfemale and 8 transmales) stopped taking GnRH analogues after a median duration of 0.8 years (range 0.1 to 3.0). Four adolescents (2.8%) discontinued GnRH analogues although they wanted to continue endocrine treatments for gender dysphoria: ▪ 1 transmale stopped due to increase in mood problems, suicidal thoughts and confusion attributed to GnRH analogues (later had gender-affirming hormones at an adult gender clinic)[1] ▪ 1 transmale experienced hot flushes, increased migraines, had a fear of injections, stress at school and unrelated medical issues, and temporarily discontinued treatment (after 4 months)[2]	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection** 1. somewhat representative 2. no-non exposed cohort 3. secure record 4. yes **Domain 2: Comparability** 1. no comparator **Domain 3: Outcome** 1. record linkage 2. yes 3. complete follow-up **Overall quality is assessed as poor.** Other comments: Physical and psychological comorbidity was poorly reported, concomitant use of other medicines was not reported. Source of funding: not reported.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
	years), respectively at commencement of GnRH analogues. Of the 143 adolescents in the study, 125 (87%, 36 transfemales and 89 transmales) subsequently started treatment with gender-affirming hormones after median 1.0 (range 0.5 to 3.8) years and 0.8 (0.3 to 3.7) years, respectively. Median age at the start of gender-affirming hormones was 16.2 years (range 14.5 to 18.6 years) in transfemales and 17.1 years (range 14.9 to 18.8 years) in transmales. Five adolescents who used GnRH analogues had not started gender-affirming hormones at the time of data collection as they were not yet eligible for this treatment due to age. At the time of data collection, they had used GnRH analogues for a median duration of 2.1 years (range 1.6 to 2.8). Tanner stage was not reported. Six adolescents had been referred to a gender clinic elsewhere for further		• 1 transmale experienced mood swings 4 months after commencing GnRH analogues. After 2.2 years he developed unexplained severe nausea and rapid weight loss and due to his general condition discontinued GnRH analogues after 2.4 years[3] • 1 transmale stopped GnRH analogues as his parents were unable to regularly collect medication from the pharmacy and take him to appointments for the injections[4] Five adolescents (3.5%) stopped treatment as they no longer wished to continue with gender-affirming treatment. • 1 adolescent had been very distressed about breast development at the start of GnRH analogues and later thought that she might want to live as a woman without breasts. She did not want to live as a boy and discontinued GnRH analogues, although dreaded breast development and menstruation. • 1 adolescent experienced concurrent psychosocial problems interfering with the exploration of gender identity and did not currently want treatment.[5] • 1 adolescent felt more in between male and female and therefore did not want to continue with GnRH analogues.[6] • 1 adolescent made a social transition while using GnRH analogues and shortly after decided to discontinue treatment.[7]	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
	treatment, including 1 who had prolonged use.		• 1 adolescent discontinued after using GnRH analogues as the treatment allowed them to feel who they were.[8]	

[1] The adolescent later indicated "I was already fully matured when I started GnRH analogues, menstruations were already suppressed by contraceptives. For me, it had no added value" (transmale, age 19 years).
[2] The adolescent restarted endocrine treatment (testosterone) 5 months later.
[3] The adolescent recovered over the next 2 years and subsequently started lynestrenol and testosterone treatment.
[4] The adolescent subsequently started lynestrenol to suppress menses, he was not yet eligible for testosterone treatment.
[5] The adolescent later reflected that "The decision to stop GnRH analogues to my mind was made by the gender team, because they did not think gender dysphoria was the right diagnosis. I do still feel like a man, but for me it is okay to be just me instead of a he or a she, so for now I do not want any further treatment" (adolescent assigned female sex at birth, age 16 years).
[6] The adolescent stated "At the moment, I feel more like 'I am' instead of 'I am a woman' or 'I am a man'" (adolescent assigned female sex at birth, age 16 years).
[7] The adolescent stated that "he had fallen in love with a girl and had never had such feelings, which made him question his gender identity. At subsequent visits, he indicated that he was happy living as a man.
[8] The adolescent stated "After using GnRH analogues for the first time, I could feel who I was without the female hormones, this gave me peace of mind to think about my future. It was an inner feeling that said I am a woman" (adolescent assigned female sex at birth, age 18 years).

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Costa R, Dunsford M, Skagerberg E, et al. (2015) Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. Journal of Sexual Medicine 12(11):2206-14. United Kingdom Prospective longitudinal observational single centre cohort study Includes participants referred to the service between 2010 and 2014.	Adolescents with gender dysphoria who completed a 6-month diagnostic process using DSM-IV-TR criteria for gender dysphoria (comprising the gender dysphoria assessment and psychological interventions) either immediately eligible for treatment with GnRH analogues or delayed eligible for treatment with GnRH analogues (received psychological support without any physical intervention). No exclusion criteria were reported. The sample consisted of 201 adolescents (sex assigned at birth male to female ratio 1:1.6) mean (±SD) age 15.52±1.41 years) from a sampling frame of	**Intervention** 101 individuals were assessed as being immediately eligible for use of GnRH analogues (no specific treatment, dose or route, or frequency of administration reported but all received psychological support). **Comparison** The analyses were between the immediately eligible and delayed eligible (n=100) adolescents.	**Critical outcomes** ***Impact on gender dysphoria*** The Utrecht gender dysphoria scale (UGDS) was used to assess adolescents' gender dysphoria related discomfort. The Cronbach's alpha (α) for the study was reported as 0.76 to 0.88, suggesting good internal consistency. UGDS was only reported once, for 160 adolescents (50 sex assigned at birth males and 110 sex assigned at birth females). The assessment time point is not reported (baseline or follow-up) and the comparison for gender related discomfort was between sex assigned at birth males and sex assigned at birth females. Sex assigned at birth males had a mean (±SD) UGDS score of 51.6 [±9.7] versus sex assigned at birth females score of 56.1 [±4.3], *t*-test 4.07; p<0.001.	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection** 1. somewhat representative 2. drawn from the same community as the exposed cohort. 3. secure record 4. no **Domain 2: Comparability** 1. partial comparator **Domain 3: Outcome** 1. independent assessment (unclear if blinded) 2. yes 3. incomplete follow-up **Overall quality is assessed as poor.**

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
	436 consecutive adolescents referred to the service between 2010 and 2014. The mean (±SD) age (n=201) at the start of GnRH analogues was 16.48 [±1.26], range 13 to 17 years. The interval from the start of the diagnostic procedure to the start of puberty suppression took approximately 1.5 years [±0.63] from baseline. None of the delayed eligible individuals received puberty suppression at the time of this study. Tanner stage was not reported.	Baseline assessment (following diagnostic procedure) was followed by follow-up at 6 months from baseline (T1), 12 months from baseline (T2) and 18 months from baseline (T3).	***Impact on mental health*** Not assessed. ***Impact on quality of life*** Not assessed. **Important outcomes** ***Psychosocial impact*** The Children's Global Assessment Scale (CGAS) was used to assess adolescents' psychosocial functioning. The CGAS was administered by psychologists, psychotherapists, and psychiatrists (intra-class correlation assessment was 0.76 ≤ Cronbach's α ≤0.94). At baseline, CGAS scores were not associated with any demographic variable, in both sex assigned at birth males and sex assigned at birth females (all p>0.1). In comparison with sex assigned at birth females, sex assigned at birth males had statistically significantly lower mean (±SD) baseline CGAS scores (55.4 [±12.7] versus 59.2 [11.8]; *t*-test 2.15; p=0.03). There was no statistically significant difference in mean (±SD) CGAS scores at baseline (T0) between immediately eligible adolescents and delayed eligible adolescents (n=201, 58.72 [±11.38] versus 56.63 [±13.14]; *t*-test 1.21; p=0.23). **Immediately eligible compared with delayed eligible participants** At follow-up, there was no statistically significant difference in mean (±SD) CGAS scores at any follow-up time point (T1, T2 or T3) between immediately	Other comments: Physical and psychological comorbidity was poorly reported, concomitant use of other medicines was not reported. Large unexplained loss to follow-up (64.7%) at T3. Source of funding: not reported.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			eligible adolescents and delayed eligible adolescents: • T1, n=201, 60.89 [±12.17] versus 60.29 [±12.81]; *t*-test 0.34; p=0.73 • T2, n=121, 64.70 [±13.34] versus 62.97 [±14.10]; *t*-test 0.69; p=0.49 • T3, n=71, 67.40 [±13.93] versus 62.53 [±13.54]; *t*-test 1.49; p=0.14. **All participants** There was a statistically significant increase in mean (±SD) CGAS scores at any follow-up time point (T1, T2 or T3) compared with baseline (T0) for the all adolescents group: • T0 (n=201) versus T1 (n=201), 57.73 [±12.27] versus 60.68 [±12.47]; *t*-test 4.87; p<0.001 • T0 (n=201) versus T2 (n=121), 57.73 [±12.27] versus 63.31 [±14.41]; *t*-test 3.70; p<0.001 • T0 (n=201) versus T3 (n=71), 57.73 [±12.27] versus 64.93 [±13.85]; *t*-test 4.11; p<0.001 There was a statistically significant increase in mean (±SD) CGAS scores when comparing the follow-up period T1 to T3 but not for the periods T1 to T2 and T2 to T3, for all adolescents: • T1 (n=201) versus T2 (n=121), 60.68 [±12.47] versus 63.31 [±14.41]; *t*-test 1.73; p<0.08 • T1 (n=201) versus T3 (n=71), 60.68 [±12.47] versus 64.93 [±13.85], *t*-test 2.40; p<0.02 • T2 (n=121) versus T3 (n=71), 63.31 [±14.41] versus 64.93 [±13.85], *t*-test 0.76; p=0.45 There were no statistically significant differences in CGAS scores between sex assigned at birth males and sex	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			assigned at birth females with gender dysphoria in all the follow-up evaluations (all p>0.1). Delayed eligible and immediately eligible adolescents with gender dysphoria were not statistically significantly different for demographic variables (all p>0.1). **Immediately eligible participants** There was a statistically significant increase in mean (±SD) CGAS scores at follow-up times T2 and T3 compared with baseline (T0) but not for T0 versus T1, for the immediately eligible adolescents: • T0 (n=101) versus T1 (n=101), 58.72 [±11.38] versus 60.89 [±12.17]; *t*-test 1.31; p=0.19 • T0 (n=101) versus T2 (n=60), 58.72 [±11.38] versus 64.70 [±13.34]; *t*-test 3.02; p=0.003 • T0 (n=101) versus T3 (n=35), 58.72 [±11.38] versus 67.40 [±13.93]; *t*-test 3.66; p<0.001 There was a statistically significant increase in mean (±SD) CGAS scores when comparing the follow-up period T1 to T3 with each other but not for the periods T1 to T2 and T2 to T3, for the immediately eligible adolescents: • T1 (n=101) versus T2 (n=60), 60.89 [±12.17] versus 64.70 [±13.34]; *t*-test 1.85; p=0.07 • T1 (n=101) versus T3 (n=35), 60.89 [±12.17] versus 67.40 [±13.93], *t*-test 2.63; p<0.001 • T2 (n=60) versus T3 (n=35), 64.70 [±13.34] versus 67.40 [±13.93], *t*-test 0.94; p=0.35 The immediately eligible adolescents had a CGAS score which was not	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			statistically significantly different compared to the sample of children/ adolescents without observed psychological /psychiatric symptoms after 12 months of puberty suppression (T3, *t*=0.01, *p*=0.99).	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
de Vries A, Steensma T, Doreleijers T, et al. (2011) Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. The Journal of Sexual Medicine 8 (8):2276-83. Netherlands Prospective longitudinal observational single centre before and after study.	The sample size was 70 adolescents receiving GnRH analogues (mean age [±SD] at assessment 13.6±1.8 years) from a sampling frame of 196 consecutive adolescents referred to the service between 2000 and 2008. Inclusion criteria were if they subsequently started gender-affirming hormones between 2003 and 2009 (mean [±SD] age at start of GnRH analogues was 14.75 [±1.92] years)[1]. No specific exclusion criteria were described. No diagnostic criteria or concomitant treatments were reported. Tanner stage of the included adolescents was not reported.	**Intervention** 70 adolescents were assessed at baseline (T0) before the start of GnRH analogues (no specific treatment, dose or route of administration reported). **Comparison** The same 70 adolescents were assessed again at follow-up (T1), shortly before starting gender-affirming hormones. Not all adolescents completed all assessments for all items[2].	**Critical outcomes** ***Impact on gender dysphoria*** Impact on gender dysphoria was assessed using the Utrecht Gender Dysphoria Scale (UGDS). • There was no statistically significant difference in UGDS scores between T0 and T1 (n=41). There was a statistically significant difference between sex assigned at birth males and sex assigned at birth females, with sex assigned at birth females reporting more gender dysphoria, *F* (*df, errdf*), *P*: 15.98 (1,39), p<0.001. ***Impact on mental health*** Depressive symptoms were assessed using the Beck Depression Inventory (BDI-II). • There was a statistically significant reduction in BDI score between T0 and T1, n=41, 8.31 [±7.12] versus 4.95 [±6.72], *F* (*df, errdf*), *P*: 9.28 (1,39), p=0.004. • There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females, *F* (*df, errdf*), *P*: 3.85 (1,39), p=0.057.	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection** 1. somewhat representative of children and adolescents who have gender dysphoria 2. no non-exposed cohort 3. no description 4. no **Domain 2: Comparability** 1. study controls for age, age at start of treatment, IQ, and parental factors **Domain 3: Outcome** 1. no description 2. no/unclear 3. complete **Overall quality is assessed as poor.** Other comments: Physical and psychological comorbidity was not reported, concomitant use of other medicines was not reported. Source of funding: This study was supported by a personal

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			Anger and anxiety were assessed using Trait Anger and Anxiety (TPI and STAI, respectively) Scales of the State-Trait Personality Inventory. • There was no statistically significant difference in anger (TPI) scale scores between T0 and T1 (n=41). There was a statistically significant difference between sex assigned at birth males and sex assigned at birth females, with sex assigned at birth females reporting increased anger compared with sex assigned at birth males, *F* (*df, errdf*), *P*: 5.70 (1,39), p=0.022. • Similarly, there was no statistically significant difference in anxiety (STAI) scale scores between T0 and T1 (n=41). There was a statistically significant difference between sex assigned at birth males and sex assigned at birth females, with sex assigned at birth females reporting increased anxiety compared with sex assigned at birth males, *F* (*df, errdf*), *P*: 16.07 (1,39), p<0.001. ***Impact on quality of life*** Not assessed. **Important outcomes** ***Impact on body image*** Impact on body image was assessed using the Body Image Scale to measure body satisfaction (BIS). There was no statistically significant difference between T0 and T1 for any of the 3 BIS scores (primary sex characteristics, secondary sex characteristics or neutral characteristics,	grant awarded to the first author by the Netherlands Organization for Health Research and Development.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			n=57). There were statistically significant differences between sex assigned at birth males and sex assigned at birth females, with sex assigned at birth females reporting more dissatisfaction, for: • primary sexual characteristics, *F* (*df, errdf*), *P*: 4.11 (1,55), p=0.047. • secondary sexual characteristics, *F* (*df, errdf*), *P*: 11.57 (1,55), p=0.001. But no statistically significant difference between sex assigned at birth males and sex assigned at birth females was found for neutral characteristics. However, there was a significant interaction effect between sex assigned at birth sex and the changes of gender dysphoria between T0 and T1; sex assigned at birth females became more dissatisfied with their secondary sex characteristics compared with sex assigned at birth males, *F* (*df, errdf*), *P*: 14.59 (1,55), p<0.001) and neutral characteristics, *F* (*df, errdf*), *P*: 15.26 (1,55), p<0.001). ***Psychosocial impact*** Psychosocial impact was assessed using both the Child Behaviour Checklist (CBCL) and the Youth Self-Report (YSR) to parents and adolescents, respectively. The Children's Global Assessment Scale was also reported. There was a statistically significant decrease in mean (±SD) total, internalising, and externalising[3] parental CBCL scores between T0 and T1[4] for all adolescents (n=54): • Total score (T0 – T1) 60.70 [±12.76] versus 54.46 [±11.23], *F* (*df, errdf*), *P*: 26.17 (1,52), p<0.001.	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			• Internalising score (T0 – T1) 61.00 [±12.21] versus 54.56 [±10.22], *F* (*df, errdf*), *P*: 22.93 (1,52), p<0.001. • Externalising score (T0 – T1) 58.04 [±12.99] versus 53.81 [±11.86], *F* (*df, errdf*), *P*: 12.04 (1,52), p=0.001. There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for total and internalising CBCL score but there was a significant difference for the externalising score: • Externalising score, *F* (*df, errdf*), *P*: 6.29 (1,52), p=0.015. There was a statistically significant decrease in mean (±SD) total, internalising, and externalising[3] YSR scores between T0 and T1 for all adolescents (n=54): • Total score (T0 – T1) 55.46 [±11.56] versus 50.00 [±10.56], *F* (*df, errdf*), *P*: 16.24 (1,52), p<0.001. • Internalising score (T0 – T1) 56.04 [±12.49] versus 49.78 [±11.63], *F* (*df, errdf*), *P*: 15.05 (1,52), p<0.001. • Externalising score (T0 – T1) 53.30 [±11.87] versus 49.98 [±9.35], *F* (*df, errdf*), *P*: 7.26 (1,52), p=0.009. There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for total and internalising YSR score but there was a significant difference for the externalising score: • Externalising score, *F* (*df, errdf*), *P*: 9.14 (1,52), p=0.004. There was a statistically significant increase in CGAS mean (±SD) score between T0 and T1 (n=41), 70.24 [±10.12] versus 73.90 [±9.63], *F* (*df, errdf*), *P*: 8.76	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			(1,39), p=0.005. There was a statistically significant difference between sex assigned at birth males and sex assigned at birth females, with sex assigned at birth females reporting lower score for global functioning compared with sex assigned at birth males, *F* (*df*, *errdf*), *P*: 5.77 (1,52), p=0.021. The proportion of adolescents scoring in the clinical range significantly decreased between T0 and T1, on the CBCL total problem scale (44.4% versus 22.2%, X^2[1] = 6.00, p=0.001), and the internalising scale (29.6% versus 11.1%, X^2[1] = 5.71, p=0.017) of the YSR.	

[1] There were statistically significant mean age [±SD] differences between sex assigned at birth males and sex assigned at birth females for age at assessment (13.14 [±1.55] versus 14.10 [±1.99] years, p=0.028), age at start of GnRH analogues (14.25 [±1.79] versus 15.21 [±1.95] years, p=0.036) and age at the start of gender-affirming hormones (16.24 [±1.21] versus 16.99 [±1.09] years, p=0.008). No statistically significant differences were seen for other baseline characteristics, time between GnRH analogue and gender-affirming hormones, full scale IQ, parental marital status, education, and sexual attraction to own, other or both sexes.
[2] Independent t-tests between mean scores on the CBCL, YSR, BDI, TPI, STAI, CGAS, UGS, and BIS of adolescents who completed both assessments and mean scores of adolescents who completed only one of the assessments revealed no significant differences on all used measures, at neither T0 or at T1.
[3] The CBCL/YSR has 2 components; Internalising score which sums the anxious/depressed, withdrawn-depressed, and somatic complaints scores; externalising score which sums rule-breaking and aggressive behaviour. The total problems score is the sum of the scores of all the problem items. The YSR is a child self-report version of the CBCL.
[4] A repeated measures ANOVA (analysis of variance) was used.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Joseph T, Ting J, Butler G. (2019) The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort. Journal of pediatric endocrinology & metabolism 32(10): 1077-1081 United Kingdom	Adolescents (12 to 14 years) with gender dysphoria (no diagnostic criteria described), n=70, including 31 transfemales and 39 transmales. All had been seen and assessed by a Gender Identity Development Service multi-disciplinary psychosocial health team for at least 4 assessments over a minimum of 6 months. All participants had entered puberty	Treatment with a GnRH analogue for at least 1 year or ongoing until they reached 16 years. No specific treatment, dose or route of administration reported. No concomitant treatments were reported.	**Critical outcomes** No critical outcomes assessed. **Important outcomes** ***Bone density: lumbar[1]*** **Lumbar spine bone mineral apparent density (BMAD)[2] 0 to 1 year** Transfemales (mean [±SD]): 0.235 (0.030) g/cm3 at baseline, 0.233 g/cm3 (0.029) at 1 year (p=0.459); z-score 0.859 (0.154) at baseline, −0.228 (1.027) at 1 year (p=0.000) Transmales (mean [±SD]):	This study was appraised using the Newcastle-Ottawa quality assessment checklist for cohort studies. **Domain 1: Selection** 1. Somewhat representative of children and adolescents who have gender dysphoria 2. Not applicable 3. Via routine clinical records 4. No

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Retrospective longitudinal observational single centre study To investigate whether there is any significant loss of bone mineral density (BMD) and bone mineral apparent density (BMAD) for up to 3 years of GnRH analogues. To investigate whether there was a significant drop after 1 year of treatment following abrupt withdrawal. 2011 to 2016	and all but 2 of the transmales were postmenarchal. 57% of the transfemales were in early puberty (G2–3 and testicular volume >4 mL) and 43% were in late puberty (G4–5). Details of the sampling frame were not reported. Further details of how the sample was drawn are not reported.	No comparator.	0.196 (0.035) g/cm3 at baseline, 0.201 (0.033) g/cm3 at 1 year (p=0.074); z-score −0.186 (1.230) at baseline, −0.541 (1.396) at 1 year (p=0.006) **Lumbar spine BMAD 0 to 2 years** Transfemales (mean [±SD]): 0.240 (0.027) g/cm3 at baseline, 0.240 (0.030) g/cm3 at 2 years (p=0.865); z-score 0.486 (0.809) at baseline, −0.279 (0.930) at 2 years (p=0.000) Transmales (mean [±SD]): 0.195 (0.058) g/cm3 at baseline, 0.198 (0.055) at 2 years (p=0.433); z-score −0.361 (1.439) at baseline, −0.913 (1.318) at 2 years (p=0.001) **Lumbar spine bone mineral density (BMD) 0 to 1 year** Transfemales (mean [±SD]): 0.860 (0.154) kg/m2 at baseline, 0.859 (0.129) kg/m2 at 1 year (p=0.962); z-score −0.016 (1.106) at baseline, −0.461 (1.121) at 1 year (p=0.003) Transmales (mean [±SD]): 0.694 (0.149) kg/m2 at baseline, 0.718 (0.124) kg/m2 at 1 year (p=0.006); z-score −0.395 (1.428) at baseline, −1.276 (1.410) at 1 year (p=0.000) **Lumbar spine BMD 0 to 2 years** Transfemales (mean [±SD]): 0.867 (0.141) kg/m2 at baseline, 0.878 (0.130) kg/m2 at 2 years (p=0.395); z-score 0.130 (0.972) at baseline, −0.890 (1.075) at 2 years (p=0.000) Transmales (mean [±SD]): 0.695 (0.220) kg/m2 at baseline, 0.731 (0.209) kg/m2 at 2 years (p=0.058); z-score −0.715 (1.406) at baseline, −2.000 (1.384) at 2 years (p=0.000) ***Bone density: femoral***	**Domain 2: Comparability** 1. No control group **Domain 3: Outcome** 1. Via routine clinical records 2. Yes 3. No statement **Overall quality is assessed as poor.** Other comments: although the evidence is of poor quality, the results suggest a possible association between GnRH analogues and BMAD. However, the results are not reliable and could be due to bias or chance. Further details of how the sample was drawn are not reported. No concomitant treatments were reported. Source of funding: None disclosed

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			Femoral neck (hip) BMD 0 to 1 year Transfemales (mean [±SD]): 0.894 (0.118) kg/m2 at baseline, 0.905 (0.104) kg/m2 at 1 year (p=0.571); z-score 0.157 (0.905) at baseline, −0.340 (0.816) at 1 year (p=0.002) Transmales (mean [±SD]): 0.772 (0.137) kg/m2 at baseline, 0.785 (0.120) kg/m2 at 1 year (p=0.797); z-score −0.863 (1.215) at baseline, −1.440 (1.075) at 1 year (p=0.000) **Femoral neck (hip) BMD 0 to 2 years** Transfemales (mean [±SD]): 0.920 (0.116) kg/m2 at baseline, 0.910 (0.125) kg/m2 at 2 years (p=0.402); z-score 0.450 (0.781) at baseline, −0.600 (1.059) at 2 years (p=0.002) Transmales (mean [±SD]): 0.766 (0.215) kg/m2 at baseline, 0.773 (0.197) at 2 years (p=0.604); z-score −1.075 (1.145) at baseline, −1.779 (0.816) at 2 years (p=0.001)	

[1] Lumbar spine (L1-L4) BMD was measured by yearly dual energy X-ray absorptiometry (DXA) scans at baseline (n=70), 1 year (n=70), and 2 years (n=31).
[2] BMAD is a size adjusted value of BMD incorporating body size measurements using UK norms in growing adolescents. Reported as g/cm3 and z-scores. Hip BMAD z-scores were not calculated as there were no available reference ranges.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Khatchadourian K, Shazhan A, Metzger D. (2014) Clinical management of youth with gender dysphoria in Vancouver. The Journal of Pediatrics 164 (4): 906-11. Canada Retrospective observational chart review single centre study	27 young people with gender dysphoria who started GnRH analogues (at mean age [±SD] 14.7±1.9 years) out of 84 young people seen at the unit between 1998 and 2011. Note: the transmale and transfemale subgroups reported in the paper is discrepant, 15 transmales and 11 transfemales (n=26) reported in the outcomes section rather than the n=27	**Intervention** 84 young people with gender dysphoria were included. For GnRH analogues no specific treatment, dose or route of administration reported. **Comparison** No comparator.	**Critical Outcomes** No critical outcomes assessed. **Important outcomes** ***Stopping treatment*** The authors report that of 15 transmales taking GnRH analogues: • 14 transitioned to testosterone treatment during the observation period • 7 continued taking GnRH analogues after starting testosterone	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection** 1. not reported 2. no non-exposed cohort 3. secure record 4. no **Domain 2: Comparability** 1. not applicable **Domain 3: Outcome**

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
	stated in the paper; complete outcome reporting is also incomplete for the transfemale group. Inclusion criteria were at least Tanner stage 2 pubertal development, previous assessment by a mental health professional and a confirmed diagnosis of gender dysphoria (diagnostic criteria not specified). No exclusion criteria are specified.		• 7 discontinued GnRH analogues after a median of 3.0 years (range 0.2 to 9.2 years), of which: ○ 5 discontinued after hysterectomy and salpingo-oophorectomy ○ 1 discontinued after 2.2 years (transitioned to gender-affirming hormone) ○ 1 discontinued after <2 months due to mood and emotional lability The authors report that of 11 transfemales taking GnRH analogues: • 5 received oestrogen treatment during the observation period • 4 continued taking GnRH analogues during oestrogen treatment • 1 discontinued GnRH analogues during oestrogen treatment (no reason reported) • 1 stopped GnRH analogues after a few months due to emotional lability • 1 stopped GnRH analogues before oestrogen treatment (the following year delayed due to heavy smoking) • 1 discontinued GnRH analogues after 13 months due to choosing not to pursue transition ***Safety*** Of the 27 patients treated with GnRH analogues: • 1 transmale participant developed sterile abscesses; they were switched from leuprolide acetate to triptorelin, and this was well tolerated. • 1 transmale participant developed leg pains and headaches on GnRH analogues, which eventually resolved without treatment.	1. record linkage 2. yes 3. in complete missing data **Overall quality is assessed as poor.** Other comments: mental health comorbidity was reported for all participants but not for the GnRH analogue cohort separately. Concomitant use of other medicines was not reported. Source of funding: No source of funding identified.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			• 1 participant gained 19 kg within 9 months of initiating GnRH analogues, although their body mass index was >85 percentile before GnRH analogues.	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Klink D, Caris M, Heijboer A et al. (2015) Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. The Journal of clinical endocrinology and metabolism 100(2): e270-5 Netherlands Retrospective longitudinal observational single centre study To assess BMD development during GnRH analogues and at age 22 years in adolescents with gender dysphoria who started treatment for gender dysphoria during adolescence. 1998 to 2012	34 adolescents (mean age ±SD 14.9±1.9 for transfemales and 15.0±2.0 for transmales at start of GnRH analogues). Participants were included if they met DSM-IV-TR criteria for gender identity disorder of adolescence and had been treated with GnRH analogues and gender-affirming hormones during their pubertal years. No concomitant treatments were reported.	The intervention was GnRH analogue monotherapy (triptorelin pamoate 3.75 mg subcutaneously every 4 weeks) followed by gender-affirming hormones from 16 years with discontinuation of GnRH analogue after gonadectomy. Median duration of GnRH analogue monotherapy in transfemales was 1.3 years (range, 0.5 to 3.8 years), and in transmales was 1.5 years (range, 0.25 to 5.2 years).	**Critical outcomes** No critical outcomes assessed. **Important outcomes** ***Bone density: lumbar*** **Lumbar spine bone mineral apparent density (BMAD)[1]** Change from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales (mean [±SD]): GnRH analogue: 0.22 (0.03) g/cm3, gender-affirming hormones: 0.22 (0.02) g/cm3 (NS); z-score GnRH analogue: −0.44 (1.10), gender-affirming hormones: −0.90 (0.80) (p=NS) Change from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales (mean [±SD]: GnRH analogue: 0.25 (0.03) g/cm3, gender-affirming hormones: 0.24 (0.02) g/cm3 (NS); z-score GnRH analogue: 0.28 (0.90), gender-affirming hormones: −0.50 (0.81) (p=0.004) **Lumbar spine bone mineral density (BMD)[1]** Change from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age	This study was appraised using the Newcastle-Ottawa quality assessment checklist for cohort studies. **Domain 1: Selection** 1. somewhat representative of children and adolescents who have gender dysphoria 2. not applicable 3. via routine clinical records 4. no **Domain 2: Comparability** 1. no control group **Domain 3: Outcome** 1. via routine clinical records 2. yes 3. follow-up rate variable across timepoints and no description of those lost **Overall quality is assessed as poor.** Other comments: Within person comparison. Small numbers of participants in each subgroup. No concomitant treatments or comorbidities were reported. Source of funding: None disclosed

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			16.6±1.4) in transfemales (mean [±SD]): GnRH analogue: 0.84 (0.13) g/m2, gender-affirming hormones: 0.84 (0.11) g/m2 (NS); z-score GnRH analogue: −0.77 (0.89), gender-affirming hormones: −1.01 (0.98) (NS) Change from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales (mean [±SD]): GnRH analogue: 0.95 (0.12) g/m2, gender-affirming hormones: 0.91 (0.10) g/m2 (p=0.006); z-score GnRH analogue: 0.17 (1.18), gender-affirming hormones: −0.72 (0.99) (p<0.001) ***Bone density; femoral*** **Femoral area BMAD[1]** Change from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales (mean [±SD]), GnRH analogue: 0.28 (0.04) g/cm3, gender-affirming hormones: 0.26 (0.04) g/cm3 (NS); z-score GnRH analogue: −0.93 (1.22), gender-affirming hormones: −1.57 (1.74) (p=NS) Change from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales (mean [±SD]), GnRH analogue: 0.32 (0.04) g/cm3, gender-affirming hormones: 0.31 (0.04) (NS); z-score GnRH analogue: 0.01 (0.70), gender-affirming hormones: −0.28 (0.74) (NS)	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			Femoral area BMD[1] Change from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales (mean [±SD]), GnRH analogue: 0.88 (0.12) g/m2, gender-affirming hormones: 0.87 (0.08) (NS); z-score GnRH analogue: −0.66 (0.77), gender-affirming hormones: −0.95 (0.63) (NS) Change from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales (mean [±SD]), GnRH analogue: 0.92 (0.10) g/m2, gender-affirming hormones: 0.88 (0.09) (p=0.005); z-score GnRH analogue: 0.36 (0.88), gender-affirming hormones: −0.35 (0.79) (p=0.001)	

[1] BMD and BMAD of the lumbar spine and femoral region (nondominant side) measured by DXA scans at start of GnRH analogues, (n=32), start of gender-affirming hormones (n=34), and at 22 years (n=34).

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Schagen SEE, Cohen-Kettenis PT, Delemarre-van de Waal HA et al. (2016) Efficacy and Safety of Gonadotropin-Releasing Hormone Agonist Treatment to Suppress Puberty in Gender Dysphoric Adolescents. The journal of sexual medicine 13(7): 1125-32	Adolescents with gender dysphoria (n=116), median age (range) 13.6 years (11.6 to 17.9) in transfemales and 14.2 years (11.1 to 18.6) in transmales during first year of GnRH analogues. Participants were included if they met DSM-IV-TR criteria for gender dysphoria, had lifelong extreme gender dysphoria, were psychologically stable and were living in a supportive environment. No concomitant treatments were	GnRH analogue monotherapy (triptorelin pamoate 3.75 mg at 0, 2 and 4 weeks followed by injections every 4 weeks, route of administration not described) for at least 3 months.	**Critical outcomes** No critical outcomes assessed. **Important outcomes** ***Other safety outcomes: liver function*** Glutamyl transferase was not elevated at baseline or during treatment in any subject. Mild elevations of aspartate aminotransferase (AST) and alanine aminotransferase (ALT) above the reference range were present at baseline but were not more prevalent during treatment than at baseline. Glutamyl transferase, AST, and ALT	This study was appraised using the Newcastle-Ottawa quality assessment checklist for cohort studies. **Domain 1: Selection** 1. somewhat representative of children and adolescents who have gender dysphoria 2. not applicable 3. via routine clinical records 4. no **Domain 2: Comparability** 1. no control group

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Netherlands Prospective longitudinal study To describe the changes in Tanner stage, testicular volume, gonadotropins, and sex steroids during GnRH analogues of adolescents with gender dysphoria to evaluate the efficacy. To report on liver enzymes, renal function and changes in body composition. 1998 to 2009	reported.		levels did not significantly change from baseline to 12 months of treatment. No values or statistical analyses were reported. ***Other safety outcomes: kidney function*** **Change in serum creatinine between 0 and 1 year** Transfemales (mean [±SD]): 70 (12) micromol/l at baseline, 66 (13) micromol/l at 1 year (p=0.20) Transmales (mean [±SD]): 73 (8) micromol/l at baseline, 68 (13) micromol/l at 1 year (p=0.01)	**Domain 3: Outcome** 1. via routine clinical records 2. yes 3. no statement **Overall quality is assessed as poor.** Other comments: Within person comparison. No concomitant treatments or comorbidities were reported. Source of funding: Ferring pharmaceuticals (triptorelin manufacturer)

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Staphorsius A, Baudewijntje P, Kreukels P, et al. (2015) Puberty suppression and executive functioning: an fMRI-study in adolescents with gender dysphoria. Psychoneuroendocrinology 565:190-9. Netherlands Cross-sectional (single time point) assessment single centre study	The inclusion criteria were diagnosed with Gender Identity Disorder according to the DSM-IV-TR and at least 12 years old and Tanner stage of at least B2 or G2 to G3 with measurable oestradiol and testosterone levels in girls and boys, respectively. For all group's exclusion criteria were an insufficient command of the Dutch language (how assessed not reported), unadjusted endocrine disorders, neurological or psychiatric disorders that could lead to deviant test results (details not reported) use	**Intervention** GnRH analogues (triptorelin pamoate 3.75 mg every 4 weeks subcutaneously or intramuscularly). **Comparison** The comparison was between adolescents with gender dysphoria receiving GnRH analogues and those without GnRH	**Critical Outcomes** No critical outcomes assessed. **Important outcomes** ***Psychosocial impact*** The Child Behaviour Checklist (CBCL) was used to assess psychosocial impact. The CBCL was administered once during the study. The reported outcomes for each group were (n, mean [±SD]): • Transfemales (all, n=18) 57.8 [±9.2] • Transfemales on GnRH analogues (n=8) 57.4 [±9.8] • Transfemales without GnRH analogues (n=10) 58.2 [±9.3]	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. somewhat representative of children and adolescents who have gender dysphoria 2. drawn from the same community as the exposed cohort 3. via routine clinical records 4. no **Domain 2: Comparability** 1. study controls for age and diagnosis

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
	of psychotropic medication, and contraindications for an MRI scan. Additionally, adolescents receiving puberty delaying medication or any form of hormones besides oral contraceptives were excluded as controls. The sample size was 85 of whom 41 were adolescents (the numbers are discrepant with the number for whom outcomes are reported n=40) with gender dysphoria (20 of whom were being treated with GnRH analogues); 24 girls and 21 boys without gender dysphoria acted as controls (not further reported here). Details of the sampling frame are not reported. The ages at which GnRH analogues were started was not reported. The mean duration of treatment was 1.6 years (SD 1.0) Mean (±SD) Tanner stage for each group was reported: • Transfemales 3.9 [±1.1] • Transfemales on GnRH analogues 4.1 [±1.0] • Transfemales without GnRH analogues 3.8 [±1.1] • Transmales 4.5 [±0.9] • Transmales on GnRH analogues 4.1 [±1.1] Transmales without GnRH analogues 4.9 [±0.3]	analogues.	• Transmales (all, n=22) 60.4 [±10.2] • Transmales on GnRH analogues (n=12) 57.5 [±9.4] • Transmales without GnRH analogues (n=10) 63.9 [±10.5] The analysis of the CBCL data is not discussed, and statistical analysis is unclear. ***Cognitive development or functioning*** **IQ[1]** • Transfemales (mean [±SD]) on GnRH analogues: 94.0 (10.3) • Transfemales (mean [±SD]) without GnRH analogues: 109.4 (21.2) • Transmales (mean [±SD]) on GnRH analogues: 95.8 (15.6) • Transmales (mean [±SD]) without GnRH analogues: 98.5 (15.9) **Reaction time[2]** • Transfemales (mean [±SD]) on GnRH analogues: 10.9 (4.1) • Transfemales (mean [±SD]) without GnRH analogues: 9.9 (3.1) • Transmales (mean [±SD]) on GnRH analogues: 9.9 (3.1) • Transmales (mean [±SD]) without GnRH analogues: 10.0 (2.0) **Accuracy[3]** • Transfemales (mean [±SD]) on GnRH analogues: 73.9 (9.1) • Transfemales (mean [±SD]) without GnRH analogues: 83.4 (9.5) • Transmales (mean [±SD]) on GnRH analogues: 85.7 (10.5) • Transmales (mean [±SD]) without	**Domain 3: Outcome** 1. via clinical assessment 2. yes 3. unclear **Overall quality is assessed as poor.** Other comments: Physical and psychological comorbidity was not reported, concomitant use of other medicines was not reported. Source of funding: This work was supported by an educational grant from the pharmaceutical firm Ferring BV, and by a VICI grant (453-08-003) from the Dutch Science Foundation. The authors state that funding sources did not play a role in any component of this study.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			GnRH analogues: 88,8 (9.7)	

[1] Estimated with 4 subscales (arithmetic, vocabulary, picture arrangement, and block design) of the Wechsler Intelligence Scale for Children, third edition (WISC-III®, Wechsler 1991) or the Wechsler Adult Intelligence Scale, third edition (WAIS-III®, Wechsler 1997), depending on the participant's age.
[2] Reaction time in seconds in the Tower of London task
[3] Percentage of correct trials in the Tower of London task

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Vlot, Mariska C, Klink, Daniel T, den Heijer, Martin et al. (2017) Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone 95: 11-19 Netherlands Retrospective observational data analysis study To investigate the course of 3 bone turnover markers in relation to bonemineral density, in adolescents with gender dysphoria during GnRH analogue and gender-affirming hormones. 2001 to 2011	Adolescents with gender dysphoria, n=70. Median age (range) 15.1 years (11.7 to 18.6) for transmales and 13.5 years (11.5 to 18.3) for transfemales at start of GnRH analogues. Participants were included if they had a diagnosis of gender dysphoria according to DSM-IV-TR criteria who were treated with GnRH analogues and then gender-affirming hormones. No concomitant treatments were reported. The study categorised participants into a young and old pubertal group, based on their bone age. The young transmales had a bone age of <14 years and the old transmales had a bone age of ≥14 years. The young transfemales group had a bone age of <15 years and the old transfemales group ≥15 years.	GnRH analogues (triptorelin pamoate 3.75 mg every 4 weeks subcutaneously).	**Critical outcomes** No critical outcomes reported **Important outcomes** ***Bone density: lumbar*** **Lumbar spine bone mineral apparent density (BMAD)** Change from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of <15 years; median [range]), GnRH analogue: 0.21 (0.17 to 0.25) g/cm3, gender-affirming hormones: 0.20 (0.18 to 0.24) g/cm3 (NS); z-score GnRH analogue: −0.20 (−1.82 to 1.18), gender-affirming hormones: −1.52 (−2.36 to 0.42) (p=0.001) Change from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of ≥15; median [range]), GnRH analogue: 0.22 (0.18 to 0.25) g/cm3, gender-affirming hormones: 0.22 (0.19 to 0.24) g/cm3 (NS); z-score GnRH analogue: −1.18 (−1.78 to 1.09), gender-affirming hormones: −1.15 (−2.21 to 0.08) (p≤0.1) Change from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of <15 years; median [range]), GnRH analogue: 0.23 (0.20 to 0.29) g/cm3, gender-affirming	This study was appraised using the Newcastle-Ottawa quality assessment checklist for cohort studies. **Domain 1: Selection** 1. Somewhat representative of children and adolescents who have gender dysphoria 2. Not applicable 3. Via routine clinical records 4. No **Domain 2: Comparability** 1. No control group **Domain 3: Outcome** 1. Via routine clinical records 2. Yes 3. Follow-up rate variable across outcomes and no description of those lost **Overall quality is assessed as poor.** Other comments: Within person comparison. No concomitant treatments were reported. Source of funding: grant from Abbott diagnostics

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			hormones: 0.23 (0.19 to 0.28) g/cm3 (NS); z-score GnRH analogue: −0.05 (−0.78 to 2.94), gender-affirming hormones: −0.84 (−2.20 to 0.87) (p=0.003) Change from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of ≥15; median [range]), GnRH analogue: 0.26 (0.21 to 0.29) g/cm3, gender-affirming hormones: 0.24 (0.20 to 0.28) g/cm3 (p≤0.01); z-score GnRH analogue: 0.27 (−1.60 to 1.80), gender-affirming hormones: −0.29 (−2.28 to 0.90) (p≤ 0.0001) ***Bone density; femoral*** **Femoral neck BMAD** Change from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of <15 years; median [range]), GnRH analogue: 0.29 (0.20 to 0.33) g/cm3, gender-affirming hormones: 0.27 (0.20 to 0.33) g/cm3 (p≤0.1); z-score GnRH analogue: −0.71 (−3.35 to 0.37), gender-affirming hormones: −1.32 (−3.39 to 0.21) (p≤0.1) Change from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of ≥15; median [range]), GnRH analogue: 0.30 (0.26 to 0.36) g/cm3, gender-affirming hormones: 0.30 (0.26 to 0.34) g/cm3 (NS); z-score GnRH analogue: −0.44 (−1.37 to 0.93), gender-affirming hormones: −0.36 (−1.50 to 0.46) (NS) Change from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of <15 years; median [range]),	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			GnRH analogue: 0.31 (0.26 to 0.36) g/cm3, gender-affirming hormones: 0.30 (0.22 to 0.35) g/cm3 (NS); z-score GnRH analogue: −0.01 (−1.30 to 0.91), gender-affirming hormones: −0.37 (−2.28 to 0.47) (NS) Change from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of ≥15; median [range]), GnRH analogue: 0.33 (0.25 to 0.39) g/cm3, gender-affirming hormones: 0.30 (0.23 to 0.41) g/cm3 (p≤0.01); z-score GnRH analogue: 0.27 (−1.39 to 1.32), gender-affirming hormones: −0.27 (−1.91 to 1.29) (p=0.002)	

Appendix F Quality appraisal checklists

Newcastle-Ottawa tool for cohort studies

Question	
Domain: Selection	
1. Representativeness of the exposed cohort	Truly representative of the average [describe] in the community Somewhat representative of the average [describe] in the community Selected group of users e.g. nurses, volunteers No description of the derivation of the cohort
2. Selection of the non-exposed cohort	Drawn from the same community as the exposed cohort Drawn from a different source No description of the derivation of the non-exposed cohort
3. Ascertainment of exposure	Secure record (e.g. surgical records) Structured interview Written self-report No description
4. Demonstration that outcome of interest was not present at start of study	Yes / No
Domain: Comparability	
1. Comparability of cohorts on the basis of the design or analysis	Study controls for [select most important factor] Study controls for any additional factor [this criteria could be modified to indicate specific control for a second important factor]
Domain: Outcome	
1. Assessment of outcome	Independent blind assessment Record linkage Self-report No description
2. Was follow-up long enough for outcomes to occur	Yes [select and adequate follow up period for outcome of interest] No
3. Adequacy of follow up of cohorts	Complete follow up (all subjects accounted for) Subjects lost to follow up unlikely to introduce bias (small number lost to follow up [select an adequate %] follow up or description provided of those lost) Follow up rate [select an adequate %] and no description of those lost No statement

Appendix G Grade profiles

Table 2: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – gender dysphoria

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on gender dysphoria									
Mean±SD Utrecht Gender Dysphoria Scale[1] (version(s) not reported), time point at baseline (before GnRH analogues) versus follow-up (before gender-affirming hormones, higher scores indicate more gender dysphoria)									
1 cohort study de Vries et al 2011	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=41	None	Baseline: 53.20±7.91 GnRH analogue: 53.9±17.42 *P*=0.333	Critical	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; *P*, P-value; SD, Standard deviation.

1 The UGDS is a validated screening tool for both adolescents and adults to assess gender dysphoria. It consists of 12 items, to be answered on a 1- to 5-point scale, resulting in a sum score between 12 and 60. The higher the UGDS score the greater the gender dysphoria.
2 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).

Table 3: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – mental health

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on mental health									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Mean±SD Beck Depression Inventory-II, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones). (Lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=41	None	Baseline: 8.31±7.12 GnRH analogue: 4.95±6.72 *P*=0.004	Critical	VERY LOW
Mean±SD Trait Anger (TPI), time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=41	None	Baseline: 18.29±5.54 GnRH analogue: 17.88±5.24 *P*=0.503	Critical	VERY LOW
Mean±SD Trait Anxiety (STAI), time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=41	None	Baseline: 39.43±10.07 GnRH analogue: 37.95±9.38 *P*=0.276	Critical	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).

Table 4: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – body image

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on body image									
Mean±SD Body Image Scale (primary sexual characteristics), time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=57	None	Baseline: 4.10±0.56 GnRH analogue: 3.98±0.71 *P*=0.145	Important	VERY LOW
Mean±SD Body Image Scale (secondary sexual characteristics), time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=57	None	Baseline: 2.74±0.65 GnRH analogue: 2.82±0.68 *P*=0.569	Important	VERY LOW
Mean±SD Body Image Scale (neutral characteristics), time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=57	None	Baseline: 2.41±0.63 GnRH analogue: 2.47±0.56 *P*=0.620	Important	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).

Table 5: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – psychosocial impact

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Psychosocial impact									
Mean [±SD] Children's Global Assessment Scale score, at baseline, higher scores indicate benefit)									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	n=101 58.72 [±11.38]	n=100 56.63 [±13.14]	*P*=0.23	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, at 6 months[2] (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	n=101 60.89 [±12.17]	n=100 60.29 [±12.81]	*P*=0.73	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, at 12 months[3] (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	n=60 64.70 [±13.34]	n=61 62.97 [±14.10]	*P*=0.49	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, at 18 months[4] (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	n=35 67.40 [±13.93]	n=36 62.53 [±13.54]	*P*=0.14	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 6 months compared to baseline (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=101 N=101	None	Baseline: 58.72±11.38 6 months: 60.89±12.17 *P*=0.19	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 12 months compared to baseline (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=101 N=60	None	Baseline: 58.72±11.38 12 months: 64.70±13.34 *P*=0.003	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Mean [±SD] Children's Global Assessment Scale score, participants at 18 months compared to baseline (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=101 N=35	None	Baseline: 58.72±11.38 18 months: 67.40±13.93 *P*<0.001	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 12 months compared to 6 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=101 N=60	None	6 months: 60.89±12.17 12 months: 64.70±13.34 *P*=0.07	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 18 months compared to 6 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=101 N=35	None	6 months: 60.89±12.17 18 months: 67.40±13.93 *P*<0.001	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 18 months compared to 12 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=60 N=35	None	12 months: 64.70±13.34 18 months: 67.40±13.93 *P*=0.35	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 6 months[2] compared to baseline (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=201	None	Baseline: 57.73±12.27 6 months: 60.68±12.47 *P*<0.001	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 12 months[3] compared to baseline (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=201 N=121	None	Baseline: 57.73±12.27 12 months: 63.31±14.41 *P*<0.001	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Mean±SD Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 18 months[4] compared to baseline (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=201 N=71	None	Baseline: 57.73±12.27 18 months: 64.93±13.85 *P*<0.001	Important	VERY LOW
Mean±SD Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 12 months compared to 6 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=201 N=121	None	6 months: 60.68±12.47 12 months: 63.31±14.41 *P*<0.08	Important	VERY LOW
Mean±SD Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 18 months compared to 6 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=201 N=71	None	6 months: 60.68±12.47 18 months: 64.93±13.85 *P*<0.02	Important	VERY LOW
Mean±SD Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 18 months compared to 12 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=121 N=71	None	12 months: 63.31±14.41 18 months: 64.93±13.85 *P*<0.45	Important	VERY LOW
Mean±SD Children's Global Assessment Scale score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, higher scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=41	None	Baseline: 70.24±10.12 GnRH analogue: 73.90±9.63 *P*=0.005	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Mean±SD Child Behaviour Checklist (total T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 60.70±12.76 GnRH analogue: 54.46±11.23 *P*<0.001	Important	VERY LOW
Mean±SD Child Behaviour Checklist (internalising T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 61.00±12.21 GnRH analogue: 52.1±9.81 *P*<0.001	Important	VERY LOW
Mean±SD Child Behaviour Checklist (externalising T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 58.04±12.99 GnRH analogue: 53.81±11.86 *P*=0.001	Important	VERY LOW
Proportion of adolescents scoring in the clinical range Child Behaviour Checklist total problem scale, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 44.4% GnRH analogue: 22.2% *P*=0.001	Important	VERY LOW
Mean±SD Youth Self-Report (total T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormone, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 55.46±11.56 GnRH analogue: 50.00±10.56 *P*<0.001	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Mean±SD Youth Self-Report (internalising T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 56.04±12.49 GnRH analogue: 49.78±11.63 *P*<0.001	Important	VERY LOW
Mean±SD Youth Self-Report (externalising T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 53.30±11.87 GnRH analogue: 49.98±9.35 *P*=0.009	Important	VERY LOW
Proportion of adolescents scoring in the clinical range Youth Self-Report (internalising T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 29.6% GnRH analogue: 11.1% *P*=0.017	Important	VERY LOW
Mean±SD Child Behaviour Checklist score, transfemales (lower scores indicate benefit									
1 cross-sectional study Staphorsius et al 2015	Serious limitations[6]	No serious indirectness	Not applicable	Not calculable	N=8	N=10	GnRH analogue: 57.4 [±9.8] No GnRH analogue: 58.2 [±9.3]	Important	VERY LOW
Mean±SD Child Behaviour Checklist score, transmales (lower scores indicate benefit)									
1 cross-sectional study Staphorsius et al 2015	Serious limitations[6]	No serious indirectness	Not applicable	Not calculable	N=12	N=10	GnRH analogues: 57.5 [±9.4] No GnRH analogue: 63.9 [±10.5]	Important	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by Costa et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).

2 6 months from baseline (after 6 months of psychological support – both groups)
3 12 months from baseline (delayed eligible gender dysphoria [GD] adolescents, after 12 months of psychological support; immediately eligible GD adolescents, after 12 months of psychological support + 6 months of puberty suppression).
4 18 months from baseline (delayed eligible gender dysphoria [GD] adolescents, after 12 months of psychological support; immediately eligible GD adolescents, after 12 months of psychological support + 6 months of puberty suppression).
5 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
6 Downgraded 1 level - the cohort study by Staphorsius et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding and no randomisation).

Table 6: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – engagement with healthcare services

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Engagement with healthcare services									
Number (proportion) failing to engage with health care services (did not attend clinic), at (up to) 9 years follow-up									
1 cohort study Brik et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	9/214 (4.2%)	None	9 adolescents out of 214 failed to attend clinic and were excluded from the study (4.2%)	Important	VERY LOW
Loss to follow-up									
1 cohort study Costa et al 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	201	None	The sample size at baseline and 6 months was 201, which dropped by 39.8% to 121 after 12 months and by 64.7% to 71 at 18 months follow-up. No explanation of the reasons for loss to follow-up are reported.	Important	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone.

1 Downgraded 1 level - the cohort study by Brik et al. (2018) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
2 Downgraded 1 level - the cohort study by Costa et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).

Table 7: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – stopping treatment

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Stopping treatment									
Number (proportion) stopping GnRH analogues, at (up to) 9 years follow-up									
1 cohort study Brik et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	9/143 (6.2%)	None	9/143 adolescents stopped GnRH analogues (6.2%)[2]	Important	VERY LOW
Number (proportion) stopping from GnRH analogues, at (up to) 13 years follow-up									
1 cohort study Khatchadourian et al 2014	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	11/27 (42%)	None	11/26 stopped GnRH analogues (42%)[4]	Important	VERY LOW
Number (proportion) stopping GnRH analogues but who wished to continue endocrine treatment, at (up to) 9 years follow-up									
1 cohort study Brik et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	4/143 (2.8%)	None	4/143 adolescents stopped GnRH analogues but wished to continue treatment (2.8%)	Important	VERY LOW
Number (proportion) stopping GnRH analogues who no longer wished gender-affirming treatment, at (up to) 9 years follow-up									
1 cohort study Brik et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	5/143 (3.5%)	None	5/143 adolescents stopped GnRH analogues and no longer wished to continue gender-affirming treatment (3.5%)	Important	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone.

1 Downgraded 1 level - the cohort study by Brik et al. (2018) was assessed as at high risk of bias (poor quality overall, lack of blinding and no control group).

2 Median duration of 0.8 years (range 0.1 to 3.0). Five adolescents stopped treatment because they no longer wished to receive gender-affirming treatment for various reasons. In 4 adolescents (all transmales), although they wanted to continue treatments for gender dysphoria, GnRH analogues were stopped mainly because of adverse effects (such as mood and emotional lability).
3 Downgraded 1 level - the cohort study by Khatchadourian et al. (2014) was assessed as at high risk of bias (poor quality overall: lack of blinding, no control group and high number of participants lost to follow-up).
4 Because of transitioning to gender-affirming hormones or gender-affirming surgery, adverse effects (such as mood and emotional lability) or no longer wishing to pursue transition.

Table 8. Question 2. For children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – bone density

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Bone density: change in lumbar BMAD									
Change in lumbar spine BMAD from baseline to 1 year in transfemales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=31	None	Mean (SD), g/cm^3 Baseline: 0.235 (0.030) 1 year: 0.233 (0.029) p=0.459 z-score Baseline: 0.859 (0.154) 1 year: −0.228 (1.027) p=0.000	IMPORTANT	VERY LOW
Change in lumbar spine BMAD from baseline to 1 year in transmales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=39	None	Mean (SD), g/cm^3 Baseline: 0.196 (0.035) 1 year: 0.201 (0.033) p=0.074 z-score Baseline: −0.186 (1.230) 1 year: −0.541 (1.396) p=0.006	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Change in lumbar spine BMAD from baseline to 2 years in transfemales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=10	None	Mean (SD), g/cm^3 Baseline: 0.240 (0.027) 2 years: 0.240 (0.030) p=0.865 z-score Baseline: 0.486 (0.809) 2 years: −0.279 (0.930) p=0.000	IMPORTANT	VERY LOW
Change in lumbar spine BMAD from baseline to 2 years in transmales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=21	None	Mean (SD), g/cm^3 Baseline: 0.195 (0.058) 2 years: 0.198 (0.055) p=0.433 z-score Baseline: −0.361 (1.439) 2 years: −0.913 (1.318) p=0.001	IMPORTANT	VERY LOW
Change in lumbar BMAD from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=11 N=12	None	Mean (SD), g/cm^3 GnRH analogue: 0.22 (0.03) Gender-affirming hormones: 0.22 (0.02) NS z-score GnRH analogue: −0.44 (1.10) Gender-affirming hormones: −0.90 (0.80) p-value: NS	IMPORTANT	VERY LOW
Change in lumbar BMAD from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=18	None	Mean (SD), g/cm^3 GnRH analogue: 0.25 (0.03) Gender-affirming hormones: 0.24 (0.02) NS z-score GnRH analogue: 0.28 (0.90) Gender-affirming hormones: −0.50 (0.81) p-value: 0.004	IMPORTANT	VERY LOW
Change in lumbar BMAD from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of <15 years)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=15	None	Median (range), g/cm^3 GnRH analogue: 0.21 (0.17 to 0.25) Gender-affirming hormones: 0.20 (0.18 to 0.24) NS z-score GnRH analogue: −0.20 (−1.82 to 1.18) Gender-affirming hormones: −1.52 (−2.36 to 0.42) p-value: <0.01	IMPORTANT	VERY LOW
Change in lumbar BMAD from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of ≥15)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=5	None	Median (range), g/cm^3 GnRH analogue: 0.22 (0.18 to 0.25) Gender-affirming hormones: 0.22 (0.19 to 0.24) NS z-score GnRH analogue: −1.18 (−1.78 to 1.09)	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							Gender-affirming hormones: −1.15 (−2.21 to 0.08) p-value: p≤0.1		
Change in lumbar BMAD from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of <14 years)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=11	None	Median (range), g/cm^3 GnRH analogue: 0.23 (0.20 to 0.29) Gender-affirming hormones: 0.23 (0.19 to 0.28) NS z-score GnRH analogue: −0.05 (−0.78 to 2.94) Gender-affirming hormones: −0.84 (−2.20 to 0.87) p-value: ≤0.01	IMPORTANT	VERY LOW
Change in lumbar BMAD from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of ≥14)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=23	None	Median (range), g/cm3 GnRH analogue: 0.26 (0.21 to 0.29) Gender-affirming hormones: 0.24 (0.20 to 0.28) p≤0.01 z-score GnRH analogue: 0.27 (−1.60 to 1.80) Gender-affirming hormones: −0.29 (−2.28 to 0.90) p-value: p ≤ 0.01)	IMPORTANT	VERY LOW
Bone density: change in lumbar BMD									
Change in lumbar spine BMD from baseline to 1 year in transfemales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=31	None	Mean (SD), kg/m2 Baseline: 0.860 (0.154) 1 year: 0.859 (0.129) p=0.962 z-score Baseline: −0.016 (1.106) 1 year: −0.461 (1.121) p=0.003	IMPORTANT	VERY LOW
Change in lumbar spine BMD from baseline to 1 year in transmales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=39	None	Mean (SD), kg/m2 Baseline: 0.694 (0.149) 1 year: 0.718 (0.124) p=0.006 z-score Baseline: −0.395 (1.428) 1 year: −1.276 (1.410) p=0.000	IMPORTANT	VERY LOW
Change in lumbar spine BMD from baseline to 2 years in transfemales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=10	None	Mean (SD), kg/m2 Baseline: 0.867 (0.141) 2 years: 0.878 (0.130) p=0.395 z-score Baseline: 0.130 (0.972) 2 years: −0.890 (1.075) p=0.000	IMPORTANT	VERY LOW
Change in lumbar spine BMD from baseline to 2 years in transmales									
1 observational study	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=21	None	Mean (SD), kg/m2 Baseline: 0.695 (0.220) 2 years: 0.731 (0.209) p=0.058	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Joseph et al. (2019)							z-score Baseline: −0.715 (1.406) 2 years: −2.000 (1.384) p=0.000		
Change in lumbar BMD from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=12 N=11	None	Mean (SD), g/m2 GnRH analogue: 0.84 (0.13) Gender-affirming hormones: 0.84 (0.11) NS z-score GnRH analogue: −0.77 (0.89) Gender-affirming hormones: −1.01 (0.98) NS	IMPORTANT	VERY LOW
Change in lumbar BMD from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=18	None	Mean (SD), g/m2 GnRH analogue: 0.95 (0.12) Gender-affirming hormones: 0.91 (0.10) p-value: 0.006 z-score GnRH analogue: 0.17 (1.18) Gender-affirming hormones: −0.72 (0.99) p-value: <0.001	IMPORTANT	VERY LOW
Bone density: change in femoral neck (hip) BMD									
Change in femoral neck BMD from baseline to 1 year in transfemales									
1 observational study	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=31	None	Mean (SD), kg/m2 Baseline: 0.894 (0.118) 1 year: 0.905 (0.104) p=0.571	IMPORTANT	VERY LOW

Study	QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
						No of events/No of patients% (n/N%)		Effect		
	Risk of bias	Indirectness	Inconsistency	Imprecision		Intervention	Comparator	Result		
Joseph et al. (2019)								z-score Baseline: 0.157 (0.905) 1 year: −0.340 (0.816) p=0.002		
Change from baseline to 1 year in femoral neck BMD in transmales										
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable		N=39	None	Mean (SD), kg/m2 Baseline: 0.772 (0.137) 1 year: 0.785 (0.120) p=0.797 z-score Baseline: −0.863 (1.215) 1 year: −1.440 (1.075) p=0.000	IMPORTANT	VERY LOW
Change from baseline to 2 years in femoral neck BMD in transfemales										
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable		N=10	None	Mean (SD), kg/m2 Baseline: 0.920 (0.116) 2 years: 0.910 (0.125) p=0.402 z-score Baseline: 0.450 (0.781) 2 years: −0.600 (1.059) p=0.002	IMPORTANT	VERY LOW
Change from baseline to 2 years in femoral neck BMD in transmales										
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable		N=21	None	Mean (SD), kg/m2 Baseline: 0.766 (0.215) 2 years: 0.773 (0.197) p=0.604 z-score Baseline: −1.075 (1.145) 2 years: −1.779 (0.816) p=0.001	IMPORTANT	VERY LOW

This document was prepared in October 2020

Page 115 of 131

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Bone density: change in femoral neck (hip) BMAD									
Change from starting GnRH analogue to starting gender-affirming hormones in femoral neck BMAD in transfemales (bone age of <15 years)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=16	None	Median (range), g/cm3 GnRH analogue: 0.29 (0.20 to 0.33) Gender-affirming hormones: 0.27 (0.20 to 0.33) p≤0.1 z-score GnRH analogue: −0.71 (−3.35 to 0.37) Gender-affirming hormones: −1.32 (−3.39 to 0.21) p≤0.1	IMPORTANT	VERY LOW
Change in femoral neck BMAD from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of ≥15)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=6	None	Median (range), g/cm3 GnRH analogue: 0.30 (0.26 to 0.36) Gender-affirming hormones: 0.30 (0.26 to 0.34) NS z-score GnRH analogue: −0.44 (−1.37 to 0.93) Gender-affirming hormones: −0.36 (−1.50 to 0.46) NS	IMPORTANT	VERY LOW
Change in femoral neck BMAD from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of <14 years)									
1 observational study	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=10	None	Median (range), g/cm3 GnRH analogue: 0.31 (0.26 to 0.36) Gender-affirming hormones: 0.30 (0.22 to 0.35)	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Vlot et al. 2017							NS z-score GnRH analogue: −0.01 (−1.30 to 0.91) Gender-affirming hormones: −0.37 (−2.28 to 0.47) NS		
Change in femoral neck BMAD from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of ≥14)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=23	None	Median (range), g/cm3 GnRH analogue: 0.33 (0.25 to 0.39) Gender-affirming hormones: 0.30 (0.23 to 0.41) p-value: ≤0.01 z-score GnRH analogue: 0.27 (−1.39 to 1.32) Gender-affirming hormones: −0.27 (−1.91 to 1.29) p-value: ≤0.01	IMPORTANT	VERY LOW
Bone density: change in femoral area BMD									
Change in femoral BMD from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=14 N=6	None	Mean (SD), g/m2 GnRH analogue: 0.88 (0.12) Gender-affirming hormones: 0.87 (0.08) NS z-score GnRH analogue: −0.66 (0.77) Gender-affirming hormones: −0.95 (0.63) NS	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Change in femoral BMD from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=18 N=13	None	Mean (SD), g/m2 GnRH analogue: 0.92 (0.10) Gender-affirming hormones: 0.88 (0.09) p-value: 0.005 z-score GnRH analogue: 0.36 (0.88) Gender-affirming hormones: −0.35 (0.79) p-value: 0.001	IMPORTANT	VERY LOW
Bone density: change in femoral area BMAD									
Change in femoral BMAD from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=12 N=10	None	Mean (SD), g/cm3 GnRH analogue: 0.28 (0.04) Gender-affirming hormones: 0.26 (0.04) NS z-score GnRH analogue: −0.93 (1.22) Gender-affirming hormones: −1.57 (1.74) p-value: NS	IMPORTANT	VERY LOW
Change in femoral BMAD from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=18 N=18	None	Mean (SD), g/cm3 GnRH analogue: 0.32 (0.04) Gender-affirming hormones: 0.31 (0.04) NS z-score	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							GnRH analogue: 0.01 (0.70) Gender-affirming hormones: −0.28 (0.74) NS		

Abbreviations: BMAD, bone mineral apparent density; BMD, bone mineral density; GnRH, gonadotrophin releasing hormone; NS, not significant; SD, standard deviation.

1 Downgraded 1 level - the cohort study by Joseph et al. (2019) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
2 Downgraded 1 level - the cohort study by Klink et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding, no randomisation, no control group and high number of participants lost to follow-up).
3 Downgraded 1 level - the cohort study by Vlot et al. (2017) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control).

Table 9 Question 2: For children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – cognitive development or functioning

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Cognitive development or functioning (1 cross-sectional study)									
IQ (4 subscales: arithmetic, vocabulary, picture arrangement, and block design) at a single time point between GnRH analogue treated and untreated transfemales									
1 Cross-sectional study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=8 Mean (SD) 94.0 (10.3)	N=10 Mean (SD) 109.4 (21.2)	NR	IMPORTANT	VERY LOW
IQ (4 subscales: arithmetic, vocabulary, picture arrangement, and block design) at a single time point between GnRH analogue treated and untreated transmales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 Cross-sectional study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=12 Mean (SD) 95.8 (15.6)	N=10 Mean (SD) 98.5 (15.9)	NR	IMPORTANT	VERY LOW
Reaction time at a single time point between GnRH analogue treated and untreated transfemales									
1 Cross-sectional study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=8 Mean (SD) 10.9 (4.1)	N=10 Mean (SD) 9.9 (3.1)	NR	IMPORTANT	VERY LOW
Reaction time at a single time point between GnRH analogue treated and untreated transmales									
1 Cross-sectional study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=12 Mean (SD) 9.9 (3.1)	N=10 Mean (SD) 10.0 (2.0)	NR	IMPORTANT	VERY LOW
Accuracy at a single time point between GnRH analogue treated and untreated transfemales									
1 cohort study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=8 Mean (SD) 73.9 (9.1)	N=10 Mean (SD) 83.4 (9.5)	NR	IMPORTANT	VERY LOW
Accuracy at a single time point between GnRH analogue treated and untreated transmales									
1 cohort study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=12 Mean (SD) 85.7 (10.5)	N=10 Mean (SD) 88.8 (9.7)	NR	IMPORTANT	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; NR, not reported; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by Staphorsius et al. (2015) was assessed as at high risk of bias (poor quality overall: lack of blinding and no randomisation).

Table 10: Question 2: In children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – other safety outcomes

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Other safety outcomes: change in serum creatinine									
Change in serum creatinine (micromol/l) between baseline and 1 year in transfemales									
1 observational study Schagen et al. 2016	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=28	None	Mean (SD) Baseline: 70 (12) 1 year: 66 (13) p-value: 0.20	IMPORTANT	VERY LOW
Change in serum creatinine (µmol/l) between baseline and 1 year in transmales									
1 observational study Schagen et al. 2016	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=29	None	Mean (SD) Baseline: 73 (8) 1 year: 68 (13) p-value: 0.01	IMPORTANT	VERY LOW
Other safety outcomes: liver enzymes									
Presence of elevated liver enzymes (AST, ALT, and glutamyl transferase) between baseline and during treatment									
1 observational study Schagen et al. 2016	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	39	None	Glutamyl transferase was not elevated at baseline or during treatment in any subject. Mild elevations of AST and ALT above the reference range were present at baseline but were not more prevalent during treatment than at baseline.	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							Glutamyl transferase, AST, and ALT levels did not significantly change from baseline to 12 months of treatment.		
Other safety outcomes: adverse effects									
Proportion of patients reporting adverse effects									
1 cohort study Khatchadourian et al 2014	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable[2]	27	None	3/27 adolescents[3]	Important	VERY LOW

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; GnRH, gonadotrophin releasing hormone; P, P-value; SD, standard deviation.

1 Downgraded 1 level - the cohort study by Schagen et al. (2016) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control).

2 Downgraded 1 level - the cohort study by Khatchadourian et al. (2014) was assessed as at high risk of bias (poor quality overall; lack of blinding, no control group and high number of participants lost to follow-up).

3 1 transmale developed sterile abscesses; they were switched from leuprolide acetate to triptorelin, and this was well tolerated. 1 transmale developed leg pains and headaches, which eventually resolved without treatment. 1 participant gained 19 kg within 9 months of initiating GnRH analogues.

Table 11: Question 4. From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may derive more (or less) advantage from treatment with GnRH analogues than the wider population of children and adolescents with gender dysphoria? – critical outcomes

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
Subgroups: sex assigned at birth males compared with sex assigned at birth females									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
Impact on gender dysphoria									
Mean [±SD] Utrecht Gender Dysphoria Scale (version(s) not reported), time point at baseline (before GnRHa) versus follow-up (just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 47.95 [±9.70] score at T1 49.67 [±9.47]	n-NR[2] score at T0 56.57 [±3.89] score at T1 56.62 [±4.0]	*F*-ratio 15.98 (*df, errdf*: 1,39), *P*<0.001	Critical	VERY LOW
Impact on mental health									
Mean [±SD] Beck Depression Inventory-II, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 5.71 [±4.31] score at T1 3.50 [±4.58]	n-NR[2] score at T0 10.34 [±8.24] score at T1 6.09 [±7.93]	*F*-ratio 3.85 (*df, errdf*: 1,39), *P*=0.057	Critical	VERY LOW
Mean [±SD] Trait Anger (TPI), time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 5.22 [±2.76] score at T1 5.00 [±3.07]	n-NR[2] score at T0 6.43 [±2.78] score at T1 6.39 [±2.59]	*F*-ratio 5.70 (*df, errdf*: 1,39), *P*=0.022	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
Mean [±SD] Trait Anxiety (STAI), time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 4.33 [±2.68] score at T1 4.39 [±2.64]	n-NR[2] score at T0 7.00 [±2.36] score at T1 6.17 [±2.69]	*F*-ratio 16.07 (*df, errdf*. 1,39), *P*<0.001	Critical	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; NR, not reported; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
2 The overall sample size completing the outcome at both time points was 41.

Table 11: Question: 4. From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may derive more (or less) advantage from treatment with GnRH analogues than the wider population of children and adolescents with gender dysphoria? – important outcomes

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
Subgroups: sex assigned at birth males compared with sex assigned at birth females									
Impact on body image									
Mean [±SD] Body Image Scale (primary sexual characteristics), time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 4.02 [±0.16] score at T1 3.74 [±0.78]	n-NR[2] score at T0 4.16 [±0.52] score at T1 4.17 [±0.58]	*F*-ratio 4.11 (*df, errdf*: 1,55), *P*=0.047	Important	VERY LOW
Mean [±SD] Body Image Scale (secondary sexual characteristics), time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 2.66 [±0.50] score at T1 2.39 [±0.69]	n-NR[2] score at T0 2.81 [±0.76] score at T1 3.18 [±0.42]	*F*-ratio 11.57 (*df, errdf*: 1,55), *P*=0.001[3]	Important	VERY LOW
Mean [±SD] Body Image Scale (neutral characteristics), time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 2.60 [±0.58] score at T1 2.32 [±0.59]	n-NR[2] score at T0 2.24 [±0.62] score at T1 2.61 [±0.50]	*F*-ratio 0.081 (*df, errdf*: 1,55), *P*=0.777[3]	Important	VERY LOW
Psychosocial impact									
Mean [±SD] Children's Global Assessment Scale score, at baseline.									
1 cohort study Costa et al 2015	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	n=not reported	n=not reported	*t*-test 2.15; *P*=0.03[5]	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
					55.4 [±12.7]	59.2 [±11.8]			
Mean [±SD] Children's Global Assessment Scale score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[6] score at T0 73.10 [±8.84] score at T1 77.33 [±8.69]	n-NR[6] score at T0 67.25 [±11.06] score at T1 70.30 [±9.44]	*F*-ratio 5.77 (*df, errdf*: 1,39), *P*=0.021	Important	VERY LOW
Mean [±SD] Child Behaviour Checklist (total T) score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 59.42 [±11.78] score at T1 50.38 [±10.57]	n-NR[7] score at T0 61.73 [±13.60] score at T1 57.73 [±10.82]	*F*-ratio 2.64 (*df, errdf*: 1,52), *P=0.110*	Important	VERY LOW
Mean [±SD] Child Behaviour Checklist (internalising T) score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 60.00 [±9.51] score at T1 52.17 [±9.81]	n-NR[7] score at T0 61.80 [±14.12] score at T1 56.30 [±10.33]	*F*-ratio 1.16 (*df, errdf*: 1,52), *P*=0.286	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
Mean [±SD] Child Behaviour Checklist (externalising T) score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 54.71 [±12.91] score at T1 48.75 [±10.22]	n-NR[7] score at T0 60.70 [±12.64] score at T1 57.87 [±11.66]	*F*-ratio 6.29 (*df, errdf*: 1,52), *P*=0.015	Important	VERY LOW
Mean [±SD] Youth Self-Report (total T) score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 53.56 [±12.26] score at T1 47.84 [±10.86]	n-NR[7] score at T0 57.10 [±10.87] score at T1 51.86 [±10.11]	*F*-ratio 1.99 (*df, errdf*: 1,52), *P*=0.164	Important	VERY LOW
Mean [±SD] Youth Self-Report (internalising T) score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 55.88 [±11.81] score at T1 49.24 [±12.24]	n-NR[7] score at T0 56.17 [±13.25] score at T1 50.24 [±11.28]	*F*-ratio 0.049 (*df, errdf*: 1,52), *P*=0.825	Important	VERY LOW
Mean [±SD] Youth Self-Report (externalising T) score, time point at baseline (T0 before GnRHa) versus follow-up (T1 just before gender-affirming hormones).									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 48.72 [±11.83] score at T1 46.52 [±9.23]	n-NR[7] score at T0 57.24 [±10.59] score at T1 52.97 [±8.51]	*F*-ratio 9.14 (*df, errdf* 1,52), *P*=0.004	Important	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; NR, not reported; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
2 The overall sample size completing the outcome at both time points was 57.
3 There was a significant interaction effect between sex assigned at birth and BDI between T0 and T1; sex assigned at birth females became more dissatisfied with their secondary F (df, errdf), P: 14.59 (1,55), P<0.001) and neutral F (df, errdf), P: 15.26 (1,55), P<0.001) sex characteristics compared with sex assigned at birth males.
4 Serious limitations – the cohort study by Costa et al. 2015 was assessed as at high risk of bias (poor quality).
5 At baseline, CGAS scores were not associated with any demographic variable, in both sex assigned at birth males and females. There were no statistically significant differences in CGAS scores between gender dysphoric sex assigned at birth males and females in all follow-up evaluations (P>0.1; full data not reported).
6 The overall sample size completing the outcome at both time points was 41
7 The overall sample size completing the outcome at both time points was 54.

Glossary

Beck Depression Inventory-II (BDI-II)	The BDI-II is a tool for assessing depressive symptoms. There are no specific scores to categorise depression severity, but it is suggested that 0 to 13 is minimal symptoms, 14 to 19 is mild depression, 20 to 28 is moderate depression, and severe depression is 29 to 63.
Body Image Scale (BIS)	The BIS is used to measure body satisfaction. The scale consists of 30 body features, which the person rates on a 5-point scale. Each of the 30 items falls into one of 3 basic groups based on its relative importance as a gender-defining body feature: primary sex characteristics, secondary sex characteristics, and neutral body characteristics. A higher score indicates more dissatisfaction.
Bone mineral apparent density (BMAD)	BMAD is a size adjusted value of bone mineral density (BMD) incorporating body size measurements using UK norms in growing adolescents.
Child Behaviour Checklist (CBCL)	CBCL is a checklist parents complete to detect emotional and behavioural problems in children and adolescents.
Children's Global Assessment Scale (CGAS)	The CGAS tool is a validated measure of global functioning on a single rating scale from 1 to 100. Lower scores indicate poorer functioning.
Gender	The roles, behaviours, activities, attributes, and opportunities that any society considers appropriate for girls and boys, and women and men.
Gender dysphoria	Discomfort or distress that is caused by a discrepancy between a person's gender identity (how they see themselves regarding their gender) and that person's sex assigned at birth (and the associated gender role, and/or primary and secondary sex characteristics).
Gonadotrophin releasing hormone (GnRH) analogues	GnRH analogues competitively block GnRH receptors to prevent the spontaneous release of 2 gonadotropin hormones, Follicular Stimulating Hormone (FSH) and Luteinising Hormone (LH) from the pituitary gland. The reduction in FSH and LH secretion reduces oestradiol secretion from the ovaries in those whose sex assigned at birth was female and testosterone secretion from the testes in those whose sex assigned at birth was male.
Sex assigned at birth	Sex assigned at birth (male or female) is a biological term and is based on genes and how external and internal sex and reproductive organs work and respond to hormones. Sex is the label that is recorded when a baby's birth is registered.
Tanner stage	Tanner staging is a scale of physical development.
Trait Anger Spielberger scales of the State-Trait Personality Inventory (TPI)	The TPI is a validated 20-item inventory tool which measures the intensity of anger as the disposition to experience angry feelings as a personality trait. Higher scores indicate greater anger.
Transgender (including transmale and transfemale)	Transgender is a term for someone whose gender identity is not congruent with their birth-registered sex. A transmale is a person who identifies as male and a transfemale is a person who identifies as female.

Utrecht Gender Dysphoria Scale (UGDS)	The UGDS is a validated screening tool for both adolescents and adults to assess gender dysphoria. It consists of 12 items, to be answered on a 1- to 5-point scale, resulting in a sum score between 12 and 60. The higher the UGDS score the greater the impact on gender dysphoria.
Youth Self-Report (YSR)	The self-administered YSR is a checklist to detect emotional and behavioural problems in children and adolescents. It is self-completed by the child or adolescent. The scales consist of a Total problems score, which is the sum of the scores of all the problem items. An internalising problem scale sums the anxious/depressed, withdrawn-depressed, and somatic complaints scores while the externalising problem scale combines rule-breaking and aggressive behaviour.

References

Included studies

- Brik T, Vrouenraets L, de Vries M et al. (2020). Trajectories of Adolescents Treated with Gonadotropin-Releasing Hormone Analogues for Gender Dysphoria. Archives of Sexual Behaviour. [Accessed 6 August 2020]
- Costa R, Dunsford M, Skagerberg E et al. (2015) Psychological Support, Puberty Suppression, and Psychosocial Functioning in Adolescents with Gender Dysphoria. Journal of Sexual Medicine. [online] Volume 12(11), Pages 2206-2214. Available at: https://doi.org/10.1111/jsm.13034 [Accessed 7 August 2020]
- de Vries A, Steensma T, Doreleijers T et al. (2011) Puberty Suppression in Adolescents with Gender Identity Disorder: A Prospective Follow-Up Study. The Journal of Sexual Medicine Volume 8, Issue 8, August, Pages 2276-2283. [Accessed 11 August 2020].
- Joseph T, Ting J, Butler G (2019) The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort. Journal of pediatric endocrinology & metabolism 32(10): 1077-1081
- Khatchadourian K, Shazhan A, Metzger D. (2014) Clinical Management of Youth with Gender Dysphoria in Vancouver. The Journal of Pediatrics. Volume 164, Issue 4, April, Pages 906-911. [Accessed 14 August 2020]
- Klink D, Caris M, Heijboer A et al. (2015) Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. The Journal of clinical endocrinology and metabolism 100(2): e270-5
- Schagen SEE, Cohen-Kettenis PT, Delemarre-van de Waal HA et al. (2016) Efficacy and Safety of Gonadotropin-Releasing Hormone Agonist Treatment to Suppress Puberty in Gender Dysphoric Adolescents. The journal of sexual medicine 13(7): 1125-32
- Staphorsius A, Baudewijntje P, Kreukels P, et al. (2015) Puberty suppression and executive functioning: An fMRI-study in adolescents with gender dysphoria. Psychoneuroendocrinology Volume 565, Pages 190-199. [Accessed 10 August 2020]

- Vlot, Mariska C, Klink, Daniel T, den Heijer, Martin et al. (2017) Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone 95: 11-19

Other references

- World Health Organisation (2018) International Classification of Diseases 11. Available from https://icd.who.int/ [online; accessed 20 August 2020]
- American Psychiatric Association. (2013). Diagnostic and statistical Manual of Mental Disorders (DSM-5) (5th ed). Washington, DC and London: American Psychiatric Publishing. pp.451-460. [accessed 20 August 2020]
- NHS England (2013). NHS Standard contract for gender identity development service for children and adolescents [accessed 20 August 2020]

Copyright

Evidence review: Gender-affirming hormones for children and adolescents with gender dysphoria

This document will help inform Dr Hilary Cass' independent review into gender identity services for children and young people. It was commissioned by NHS England and Improvement who commissioned the Cass review. It aims to assess the evidence for the clinical effectiveness, safety and cost-effectiveness of gender-affirming hormones for children and adolescents aged 18 years or under with gender dysphoria.

The document was prepared by NICE in October 2020.

The content of this evidence review was up to date on 21 October 2020. See summaries of product characteristics (SPCs), British National Formulary (BNF) or the Medicines and Healthcare products Regulatory Agency (MHRA) or NICE websites for up-to-date information.

Contents

1. Introduction

This review aims to assess the evidence for the clinical effectiveness, safety and cost-effectiveness of gender-affirming hormones for children and adolescents aged 18 years or under with gender dysphoria. The review follows the NHS England Specialised Commissioning process and template and is based on the criteria outlined in the PICO framework (see appendix A). This document will help inform Dr Hilary Cass' independent review into gender identity services for children and young people.

Gender dysphoria in children, also known as gender identity disorder or gender incongruence of childhood (World Health Organisation 2020), refers to discomfort or distress that is caused by a discrepancy between a person's gender identity (how they see themselves[1] regarding their gender) and that person's sex assigned at birth and the associated gender role, and/or primary and secondary sex characteristics (Diagnostic and Statistical Manual of Mental Disorders 2013).

Gender-affirming hormones are oestradiol for sex assigned at birth males (transfemales) and testosterone for sex assigned at birth females (transmales). The aim of gender-affirming hormones is to induce the development of the physical sex characteristics congruent with the individual's gender expression while aiming to improve mental health and quality of life outcomes.

No oestradiol-containing products are licensed for gender dysphoria and therefore any use for children and adolescents with gender dysphoria is off-label.

The only testosterone-containing product licensed for gender dysphoria is Sustanon 250 mg/ml solution for injection, which is indicated as supportive therapy for transmales, use of all other testosterone-containing products for children and adolescents with gender dysphoria is off-label.

For children and adolescents with gender dysphoria it is recommended that management plans are tailored to the needs of the individual and aim to ameliorate the potentially negative impact of gender dysphoria on general developmental processes, to support young people and their families in managing the uncertainties inherent in gender identity development and to provide ongoing opportunities for exploration of gender identity. The plans may also include psychological support and exploration and, for some individuals, the use of gonadotrophin releasing hormone (GnRH) analogues in adolescence to suppress puberty; this may be followed later with gender-affirming hormones of the desired sex (NHS England 2013).

Currently NHS England, as part of the Gender Identity Development Service for Children and Adolescents, routinely commissions gender-affirming hormones for young people with continuing gender dysphoria from around their 16th birthday subject to individuals meeting the eligibility and readiness criteria (Clinical Commissioning Policy 2016).

---

[1] Gender refers to the roles, behaviours, activities, attributes and opportunities that any society considers appropriate for girls and boys, and women and men (World Health Organisation, Health Topics: Gender).

2. Executive summary of the review

Ten observational studies were included in the evidence review. Seven studies were retrospective observational studies (Allen et al. 2019, Kaltiala et al. 2020, Khatchadourian et al. 2014, Klaver et Al. 2020, Klink et al. 2015, Stoffers et al. 2019, Vlot et al. 2017) and 3 studies were prospective longitudinal observational studies (Achille et al. 2020, Kuper et al. 2020, Lopez de Lara et al. 2020). No studies directly compared gender-affirming hormones to a control group (either placebo or active comparator). Follow-up was relatively short across all studies, with an average duration of treatment with gender-affirming hormones between around 1 year and 5.8 years.

The terminology used in this topic area is continually evolving and is different depending on stakeholder perspectives. In this evidence review we have used the phrase 'people's assigned sex at birth' rather than saying natal or biological sex and 'cross sex hormones' are now referred to as 'gender-affirming hormones'. The research studies may use historical terms which are no longer considered appropriate.

In children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Critical outcomes

The critical outcomes for decision making are impact on gender dysphoria, impact on mental health and quality of life. The quality of evidence for all these outcomes was assessed as very low certainty using modified GRADE.

Impact on gender dysphoria

The study by Lopez de Lara et al. 2020 in 23 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, gender dysphoria (measured using the Utrecht Gender Dysphoria Scale [UGDS]) was statistically significantly reduced (improved) from a mean [±SD] score of 57.1 (±4.1) points at baseline to 14.7 (±3.2) points at 12 months, which is below the threshold (40 points) for gender dysphoria ($p<0.001$).

Impact on mental health

Depression

The study by Lopez de Lara et al. 2020 in 23 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, depression (measured using the Beck Depression Inventory-II [BDI-II]) was statistically significantly reduced from a mean [±SD] score of 19.3 (±5.5) points at baseline to 9.7 (±3.9) points at 12 months ($p<0.001$).

The study by Achille et al. 2020 in 50 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, depression was statistically significantly reduced from baseline to about 12 months follow-up:

- The Center for Epidemiologic Studies Depression (CESD-R) improved from a mean score of 21.4 points at baseline to 13.9 points ($p<0.001$).
- The Patient Health Questionnaire (PHQ 9) Modified for Teens improved, although absolute scores were not reported numerically ($p<0.001$).

The study by Kuper et al. 2020 in 148 adolescents with gender dysphoria (of whom 123 received gender-affirming hormones) found that during treatment with gender-affirming hormones for an average of 10.9 months, the impact on depression (measured using the Quick Inventory of Depressive Symptoms [QIDS]) was unclear as no statistical analysis was reported. The mean (±SD) self-reported score was 9.6 points (±5.0) at baseline and 7.4 (±4.5) at follow-up. The mean (±SD) clinician-reported score was 5.9 points (±4.1) at baseline and 6.0 (±3.8).

The study by Kaltiala et al. 2020 in 52 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, statistically significantly fewer participants needed treatment for depression (54% at initial assessment compared with 15% at 12-month follow-up, p<0.001). No details of the treatments for depression are reported.

Anxiety

The study by Lopez de Lara et al. 2020 in 23 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, state anxiety (measured using the State-Trait Anxiety Inventory [STAI] – State subscale) was statistically significantly reduced from a mean (±SD) score of 33.3 points (±9.1) at baseline to 16.8 points (±8.1) at 12 months (p<0.001). Trait anxiety (measured using STAI – Trait subscale) was also statistically significantly reduced from a mean (±SD) score of 33.0 (±7.2) points at baseline to 18.5 (±8.4) points at 12 months (p<0.001).

The study by Kuper et al. 2020 in 148 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, small reductions were seen in anxiety, panic, generalised anxiety, social anxiety and separation anxiety symptoms and school avoidance (measured using the Screen for Child Anxiety Related Emotional Disorders [SCARED] questionnaire) from baseline to follow-up (mean duration of treatment 10.9 months). The statistical significance of these findings are unknown as no statistical analyses were reported.

The study by Kaltiala et al. 2020 in 52 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, statistically significantly fewer participants needed treatment for anxiety (48% at initial assessment compared with 15% at 12-month follow-up, p<0.001). No details of treatments for anxiety are reported.

Suicidality and self-injury

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, suicide risk (measured using the Ask Suicide-Screening Questions [ASQ]) was statistically significantly reduced from an adjusted mean (±SE) score of 1.11 points (±0.22) at baseline to 0.27 points (±0.12) after about 12 months (p<0.001).

The study by Achille et al. 2020 in 50 adolescents with gender dysphoria (of whom 35 received gender-affirming hormones at follow-up) found that during treatment with gender-affirming hormones, the impact on suicidal ideation was unclear (measured using the PHQ 9_Modified for Teens with additional questions for suicidal ideation). At baseline 10%of participants had suicidal ideation and 6% had suicidal ideation after about 12 months, but it is unclear if these participants received gender-affirming hormones. No statistical analyses were reported.

The study by Kuper et al. 2020 in 148 adolescents with gender dysphoria reported the impact on suicidal ideation, suicide attempts and non-suicidal self-injury during treatment with gender-affirming hormones, after mean 10.9 months follow-up . The statistical significance of these findings are unknown as no statistical analyses were reported:

- Suicidal ideation was reported in 25% of participants 1 month before the initial assessment and in 38% of participants during follow-up.
- Suicide attempts were reported in 2% of participants at 3 months before the initial assessment and in 5% during follow-up.
- Self-injury was reported in 10% of participants at 3 months before the initial assessment and in 17% during follow-up.

The study by Kaltiala et al. 2020 in 52 adolescents with gender dysphoria reported that during treatment with gender-affirming hormones, statistically significantly fewer participants needed treatment for suicidal ideation or self-harm (35% at initial assessment compared with 4% at 12-month follow-up, p<0.001). No details of treatments for suicidal ideation or self-harm are reported.

Other related symptoms

The study by Kaltiala et al. 2020 in 52 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, there was no statistically significant difference in the number of people needing treatment for either psychotic symptoms or psychosis, conduct problems or antisocial behaviour, substance abuse, autism, attention deficit hyperactivity disorder (ADHD) or eating disorders during the 12-month 'real life' phase compared with before or during the assessment. No details of the treatments received are reported.

Impact on quality of life

The study by Achille et al. 2020 in 50 adolescents with gender dysphoria (of whom 35 were receiving gender-affirming hormones at follow-up) found that during treatment with gender-affirming hormones, quality of life (measured using the Quality of Life Enjoyment and Satisfaction Questionnaire [QLES-Q-SF]) was statistically significantly improved from baseline to about 12 months, but absolute scores were not reported numerically (p<0.001).

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, quality of life (measured using the General Well-Being Scale [GWBS] of the Paediatric Quality of Life Inventory) was statistically significantly improved from an adjusted mean (±SE) score of 61.70 (±2.43) points at baseline to 70.23 (±2.15) points at about 12 months (p<0.002).

Important outcomes

The important outcomes for decision making are impact on body image, psychosocial impact, engagement with healthcare services, impact on extent of and satisfaction with surgery and de-transition. The quality of evidence for all these outcomes was assessed as very low certainty using modified GRADE.

Impact on body image

The study by Kuper et al. 2020 in 148 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, the impact on body image is unclear (measured using the Body Image Scale [BIS]). The mean (±SD) BIS score was 70.7 points (±15.2) at baseline and 51.4 points (±18.3) at follow-up (mean duration of treatment 10.9 months; no statistical analysis was reported).

Psychosocial impact

The study by Lopez de Lara et al. 2020 in 23 adolescents with gender dysphoria found that during treatment with gender affirming hormones, family functioning is unchanged (measured using the Family Adaptability, Partnership, Growth, Affection and Resolve [APGAR] test). The mean score was 17.9 points at baseline and 18.0 points at 12-month follow-up (no statistical analysis was reported).

The study by Lopez de Lara et al. 2020 in 23 adolescents with gender dysphoria found that during treatment with gender affirming hormones, behavioural problems (measured using the Strengths and Difficulties Questionnaire [SDQ]) were statistically significantly improved from a mean (±SD) of 14.7 (±3.3) points at baseline to 10.3 points (±2.9) at 12-month follow-up ($p<0.001$).

The study by Kaltiala et al. 2020 in 52 adolescents with gender dysphoria found that about 12-months after starting treatment with gender-affirming hormones:

- Statistically significantly fewer participants were living with parents or guardians (73% versus 40%, $p=0.001$) and statistically significantly fewer participants had normal peer contacts (89% versus 81%, $p<0.001$).
- There were no statistically significant differences in:
  - progress in school or work (64% versus 60%, $p=0.69$),
  - the number of participants who had been dating or in steady relationships (62% versus 58%, $p=0.51$)
  - the ability to cope with matters outside of the home (for example, shopping and travelling alone on local public transport; 81% versus 81%, $p=1.0$)

Engagement with health care services

No evidence was identified.

Impact on extent of and satisfaction with surgery

No evidence was identified.

De-transition

No evidence was identified.

In children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Important outcomes

The important outcomes for decision making are short- and long-term safety outcomes and adverse effects. The quality of evidence for all these outcomes was assessed as very low certainty using modified GRADE.

Bone density

The study by Klink et al. 2015 in 34 adolescents with gender dysphoria (who were previously treated with a GnRH analogue) found that gender-affirming hormones may increase lumbar spine and femoral neck bone density. However, not all results are statistically significant (particularly in transfemales). Z-scores suggest the average bone density at the end of follow-up was generally lower than in the equivalent cisgender population (transfemales compared with cis-males and transmales compared with cis-females). From starting gender-affirming hormones to age 22 years:

- There was no statistically significant difference in lumbar spine bone mineral apparent density (BMAD) z-score in transfemales, but this was statistically significantly higher in transmales (z-score [±SD]: start of hormones -0.50 [±0.81], age 22 years -0.033 [±0.95], p=0.002).
- There was no statistically significant difference in lumbar spine bone mineral density (BMD) z-score in transfemales or transmales.
- Actual lumbar spine BMAD and BMD values were statistically significantly higher in transfemales and transmales.
- There was no statistically significant difference in femoral neck BMD z-score in transfemales, but this was statistically significantly higher in transmales (z-score [SD]: start of hormones -0.35 [0.79], age 22 years -0.35 [0.74], p=0.006).
- There was no statistically significant difference in actual femoral neck BMAD values in transfemales, but this was statistically significantly higher in transmales.
- Actual femoral neck BMD values were statistically significantly higher in transfemales and transmales.

The study by Vlot et al. 2017 in 70 adolescents with gender dysphoria (who were previously treated with a GnRH analogue) found that gender-affirming hormones may increase lumbar spine and femoral neck bone density. However, not all results are statistically significant. Z-scores suggest the average bone density at the end of follow-up was generally lower than the equivalent cisgender population (transfemales compared with cis-males and transmales compared with cis-females). From starting gender-affirming hormones to 24-month follow-up:

- The z-score for lumbar spine BMAD was statistically significantly higher in transfemales with a bone age of less than 15 years (z-score [range]: start of hormones -1.52 [-2.36 to 0.42], 24-month follow-up -1.10 [-2.44 to 0.69], p≤ 0.05) and 15 years and older (z-score [range]: start of hormones -1.15 [-2.21 to 0.08], 24-month follow-up -0.66 [-1.66 to 0.54], p≤ 0.05).
- The z-score for lumbar spine BMAD was statistically significantly higher in transmales with a bone age of less than 14 years (z-score [range]: start of hormones -0.84 [-2.2 to 0.87], 24-month follow-up -0.15 [-1.38 to 0.94], p≤ 0.01) and 14 years and older (z-score [range]: start of hormones -0.29 [-2.28 to 0.90], 24-month follow-up -0.06 [-1.75 to 1.61], p≤ 0.01).
- Actual lumbar spine BMAD values were statistically significantly higher in transfemales and transmales of all bone ages.
- There was no statistically significant difference in femoral neck BMAD z-score in transfemales (all bone ages).
- The z-score for femoral neck BMAD was statistically significantly higher in transmales with a bone age of less than 14 years (z-score [range]: start of hormones

-0.37 [-2.28 to 0.47], 24-month follow-up -0.37 [-2.03 to 0.85], p≤ 0.01) and 14 years and older (z-score [range]: start of hormones -0.27 [-1.91 to 1.29], 24-month follow-up 0.02 [-2.1 to 1.35], p≤0.05).

- There was no statistically significant difference in actual femoral neck BMAD values in transfemales (all bone ages), but this was statistically significantly higher in transmales (all bone ages).

The study by Stoffers et al. 2019 in 62 sex assigned at birth females (transmales) with gender dysphoria (who were previously treated with a GnRH analogue) found that during treatment with gender-affirming hormones there was no statistically significant difference in lumbar spine or femoral neck bone density (measured as BMD z-scores or actual values) from starting gender-affirming hormones to any timepoint (6, 12 and 24 months).

Change in clinical parameters

The study by Klaver et al. 2020 in 192 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, from starting treatment to age 22 years:

- Glucose levels, insulin levels and insulin resistance were largely unchanged in transfemales and transmales.
- Total cholesterol, HDL cholesterol and LDL cholesterol levels were unchanged in transfemales, and there was a statistically significant improvement in triglyceride levels.
- Total cholesterol, HDL cholesterol, LDL cholesterol and triglyceride levels significantly worsened in transmales, but mean levels were within the UK reference range at the end of treatment.
- Diastolic blood pressure was statistically significantly increased in transfemales and transmales. Systolic blood pressure was also statistically significantly increased in transmales, but not in transfemales. The absolute increases in blood pressure were small.
- Body mass index was statistically significantly increased in transfemales and transmales, although most participants were within the healthy weight range (18.5 to 24.9 kg/m).

The study by Stoffers et al. 2019 in 62 sex assigned at birth females (transmales) with gender dysphoria found that during treatment with gender affirming hormones, from starting treatment to 24-month follow-up:

- There was no statistically significant change in glycosylated haemoglobin (HbA1c).
- There was no statistically significant change in aspartate aminotransferase (AST), alanine aminotransferase (ALT) and gamma-glutamyltransferase (GCT).
- There was a statistically significant increase in alkaline phosphatase (ALP) at some timepoints, but the difference was not statistically significant by 24-months.
- There was a statistically significant increase in serum creatinine levels at all timepoints up to 24 months, but these were within the UK reference range. Serum urea levels were unchanged (follow-up duration not reported).

Treatment discontinuation and adverse effects

The study by Khatchadourian et al. 2014 in 63 adolescents (24 transfemales and 39 transmales) with gender dysphoria found that during treatment with gender affirming hormones (duration of treatment not reported):

- No participants permanently discontinued treatment.
- No transfemales temporarily discontinued treatment, but 3 transmales temporarily discontinued treatment due to mental health comorbidities (n=2) and androgenic alopecia (n=1). All 3 participants eventually resumed treatment, although timescales were not reported
- No severe complications were reported.
- No transfemales reported minor complications, but 12 transmales developed minor complications which were: severe acne (n=7), androgenic alopecia (n=1), mild dyslipidaemia (n=3) and significant mood swings (n=1).

In children and adolescents with gender dysphoria, what is the cost-effectiveness of gender-affirming hormones compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?

No cost-effectiveness evidence was found for gender-affirming hormones for children and adolescents with gender dysphoria.

From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria?

Some studies reported data separately for the following subgroups of children and adolescents with gender dysphoria:

- Sex assigned at birth males (transfemales)
- Sex assigned at birth females (transmales)
- Tanner stage at which GnRH analogue or gender-affirming hormones started.
- Diagnosis of a mental health condition.

Some direct comparisons of transfemales and transmales were included. No evidence was found for other specified subgroups.

Sex assigned at birth males (transfemales)

Impact on mental health

In the study by Kuper et al. 2020 in 33 to 45 (number varies by outcome) sex assigned at birth males (transfemales) with gender dysphoria found that during treatment with gender-affirming hormones changes were seen in depression, anxiety and anxiety-related symptoms from baseline to follow-up (mean duration of treatment 10.9 months). The authors did not report any statistical analyses, so it is unclear if any changes were statistically significant.

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, suicide risk (measured using the ASQ) is not statistically significant different in transfemales compared with transmales, between baseline and the final assessment at about 12 months (p=0.79).

The study by Achille et al. 2020 in 17 transfemales with gender dysphoria found that during treatment with gender-affirming hormones, suicidal ideation (measured using the PHQ 9_Modified for Teens with additional questions for suicidal ideation) was reported in 11.8%

(2/17) of transfemales at baseline compared with 5.9% (1/17) at about 12-months follow-up (no statistical analysis was reported).

Impact on quality of life

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, quality of life (measured using the GWBS of the Paediatric Quality of Life Inventory) was not statistically significant different in transfemales compared with transmales, between baseline and the final assessment at about 12 months (p=0.32).

Bone density

The studies by Klink et al. 2015 and Vlot et al. 2017 provided evidence on bone density in transfemales; see above for details.

Change in clinical parameters

The study by Klaver et al. 2020 provided evidence on the following clinical parameters in transfemales:

- Glucose levels, insulin levels and insulin resistance.
- Total cholesterol, HDL cholesterol and LDL cholesterol and triglycerides.
- Blood pressure.
- Body mass index.

See above for details.

Treatment discontinuation and adverse effects

The study by Khatchadourian et al. 2014 provided evidence on treatment discontinuation and adverse effects in transfemales; see above for details.

Sex assigned at birth females (transmales)

Impact on mental health

In the study by Kuper et al. 2020 in 65 to 78 (number varies by outcome) sex assigned at birth females (transmales) with gender dysphoria found that during treatment with gender-affirming hormones, changes were seen in depression, anxiety and anxiety-related symptoms from baseline to 10.9 month follow-up. The authors did not report any statistical analyses, so it is unclear if any changes were statistically significant.

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, suicide risk (measured using the ASQ) is not statistically significantly different in transmales compared with transfemales, between baseline and the final assessment (p=0.79).

The study by Achille et al. 2020 in 33 transmales with gender dysphoria found that during treatment with gender-affirming hormones, suicidal ideation (measured using the PHQ 9_Modified for Teens with additional questions for suicidal ideation) was reported in 9.1% (3/33) of transmales at baseline compared with 6.1% (2/33) at about 12-months follow-up (no statistical analysis reported).

Impact on quality of life

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, quality of life (measured using the GWBS of the

Paediatric Quality of Life Inventory) was not statistically significantly different in transmales compared with transfemales, between baseline and the final assessment at about 12 months (p=0.32).

Bone density
The studies by Klink et al. 2015, Stoffers et al. 2019 and Vlot et al. 2017 provided evidence on bone density in transmales; see above for details.

Change in clinical parameters
The study by Klaver et al. 2020 provided evidence on the following clinical parameters in transmales:

- Glucose levels, insulin levels and insulin resistance.
- Total cholesterol, HDL cholesterol and LDL cholesterol and triglycerides.
- Blood pressure.
- Body mass index.

See above for details.

The study by Stoffers et al. 2019 provided evidence on HbA1c, liver enzymes and renal function in transmales; see above for details.

Treatment discontinuation and adverse effects
The study by Khatchadourian et al. 2014 provided evidence on treatment discontinuation and adverse effects in transmales; see above for details.

Tanner stage at which GnRH analogues or gender-affirming hormones started
The study by Kuper et al. 2020 stated that the impact of Tanner stage on outcomes was considered, but it is unclear if this refers to Tanner stage at the initial assessment, at the start of GnRH analogue treatment or another timepoint. No results were reported.

Diagnosis of a mental health condition
Impact on mental health
The study by Achille et al. 2020 in 50 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, there was no statistically significant difference in depression (measured using the CESD-R and PHQ 9_Modified for Teens) when the results were adjusted for engagement in counselling and medicines for mental health problems, from baseline to about 12-months follow-up.

Impact on quality of life
The study by Achille et al. 2020 in 50 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, there was no statistically significant difference in quality of life (measured using the QLES-Q-SF) when the results were adjusted for engagement in counselling and medicines for mental health problems, from baseline to about 12-months follow-up.

From the evidence selected,

(a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?

(b) what were the ages at which participants commenced treatment with gender-affirming hormones?

(c) what was the duration of treatment with GnRH analogues?

The most commonly reported diagnostic criteria for gender dysphoria was the DSM criteria in use at the time (5/10 studies). In 3 studies (Klaver et al. 2020, Klink et al. 2015 and Vlot et al. 2017) DSM-IV-TR criteria was used. In 2 studies (Kuper et al. 2020 and Stoffers et al. 2019) DSM-V criteria was used. One study from Finland (Kaltiala et al. 2020) used the ICD-10 diagnosis of 'transexualism'. It was not reported how gender dysphoria was defined in the remaining 4 studies.

In the studies, treatment with gender-affirming hormones started at about 16 to 17 years, with a range of about 14 to 19 years. Most studies did not report the duration of treatment with GnRH analogues, but where this was reported there was a wide variation ranging from a few months up to about 5 years (Klaver et al. 2020, Klink et al. 2015 and Stoffers et al. 2019).

Discussion

The key limitation to identifying the effectiveness and safety of gender-affirming hormones for children and adolescents with gender dysphoria is the lack of reliable comparative studies.

All the studies included in the evidence review are uncontrolled observational studies, which are subject to bias and confounding and were of very low certainty using modified GRADE. A fundamental limitation of all the uncontrolled studies included in this review is that any changes in scores from baseline to follow-up could be attributed to a regression-to-the-mean.

The included studies have relatively short follow-up, with an average duration of treatment with gender-affirming hormones between around 1 year and 5.8 years. Further studies with a longer follow-up are needed to determine the long-term effect of gender-affirming hormones for children and adolescents with gender dysphoria.

Most studies included in this review did not report comorbidities (physical or mental health) and no study reported concomitant treatments in detail. Because of this it is not clear whether any changes seen were due to gender-affirming hormones or other treatments the participants may have received.

There is a degree of indirectness in some studies, with some participants included that fall outside of the population of this evidence review. Furthermore, participant numbers are poorly reported in some studies, with high numbers lost to follow-up or outcomes not reported for some participants. The authors provide no explanation for this incomplete reporting.

Details of the gender-affirming hormone treatment regimen are poorly reported in most of the included studies, with limited information provided about the medicines, doses and routes of administration used. It is not clear whether the interventions used in the studies are reflective of current UK practice for children and adolescents with gender dysphoria.

It is difficult to draw firm conclusions for many of the effectiveness and safety outcomes reported in the included studies because many different scoring tools and methods were used to assess the same outcome, often with conflicting results. In addition to this, most

outcomes reported across the included studies do not have an accepted minimal clinically important difference (MCID), making it difficult the determine whether any statistically significant changes seen are clinically meaningful. However, the authors of some studies report thresholds to interpret the results of the scoring tools (for example, by linking scores to symptom severity), so some conclusions can be made.

Conclusion

Any potential benefits of gender-affirming hormones must be weighed against the largely unknown long-term safety profile of these treatments in children and adolescents with gender dysphoria.

Results from 5 uncontrolled, observational studies suggest that, in children and adolescents with gender dysphoria, gender-affirming hormones are likely to improve symptoms of gender dysphoria, and may also improve depression, anxiety, quality of life, suicidality, and psychosocial functioning. The impact of treatment on body image is unclear. All results were of very low certainty using modified GRADE.

Safety outcomes were reported in 5 observational studies. Statistically significant increases in some measures of bone density were seen following treatment with gender-affirming hormones, although results varied by bone region (lumber spine versus femoral neck) and by population (transfemales versus transmales). However, z-scores suggest that bone density remained lower in transfemales and transmales compared with an equivalent cisgender population. Results from 1 study of gender-affirming hormones started during adolescence reported statistically significant increases in blood pressure and body mass index, and worsening of the lipid profile (in transmales) at age 22 years, although longer term studies that report on cardiovascular event rates are required. Adverse events and discontinuation rates associated with gender-affirming hormones were only reported in 1 study, and no conclusions can be made on these outcomes.

This review did not identify sub-groups of patients who may benefit more from gender-affirming hormones.

No cost-effectiveness evidence was found to determine whether gender-affirming hormones are a cost-effective treatment for children and adolescents with gender dysphoria.

3. Methodology

Review questions

The review question(s) for this evidence review are:

1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
2. For children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

3. For children and adolescents with gender dysphoria, what is the cost-effectiveness of gender-affirming hormones compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?
4. From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria?
5. From the evidence selected,
   (a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?
   (b) what were the ages at which participants commenced treatment with gender-affirming hormones?
   (c) what was the duration of GnRH analogues treatment?

See appendix A for the full review protocol.

Review process

The methodology to undertake this review is specified by NHS England in their 'Guidance on conducting evidence reviews for Specialised Services Commissioning Products' (2020).

The searches for evidence were informed by the PICO and were conducted on 21 July 2020.

See appendix B for details of the search strategy.

Results from the literature searches were screened using their titles and abstracts for relevance against the criteria in the PICO framework. Full text references of potentially relevant evidence were obtained and reviewed to determine whether they met the inclusion criteria for this evidence review.

See appendix C for evidence selection details and appendix D for the list of studies excluded from the review and the reasons for their exclusion.

Relevant details and outcomes were extracted from the included studies and were critically appraised using a checklist appropriate to the study design. See appendix E and appendix F for individual study and checklist details.

The available evidence was assessed by outcome for certainty using modified GRADE. See appendix G for GRADE Profiles.

4. Summary of included studies

Ten observational studies were included in the evidence review. Seven studies were retrospective observational studies (Allen et al. 2019, Kaltiala et al. 2020, Khatchadourian et al. 2014, Klaver et Al. 2020, Klink et al. 2015, Stoffers et al. 2019, Vlot et al. 2017) and three

studies were prospective longitudinal observational studies (Achille et al. 2020, Kuper et al. 2020, Lopez de Lara et al. 2020).

The terminology used in this topic area is continually evolving and is different depending on stakeholder perspectives. In this evidence review we have used the phrase 'people's assigned sex at birth' rather than saying natal or biological sex and 'cross sex hormones' are now referred to as 'gender-affirming hormones'. The research studies may use historical terms which are no longer considered appropriate.

Table 1 provides a summary of these included studies and full details are given in appendix E.

Table 1 Summary of included studies

Study	Population	Intervention and comparison	Outcomes reported
Achille et al. 2020 Prospective longitudinal study Single centre, New York, United States	50 children, adolescents and young adults with gender dysphoria; 17 transfemales and 33 transmales Mean age at baseline was 16.2 years (SD 2.2)	**Intervention** Endocrine interventions (the collective term used for puberty suppression and gender-affirming hormones) were introduced as per Endocrine Society and the World Professional Association for Transgender Health (WPATH) guidelines Puberty suppression was: • GnRH analogue and/or anti-androgens (transfemales) • GnRH analogue or medroxyprogesterone (transmales) Once eligible, gender-affirming hormones were offered, these were: • Oestradiol (transfemales) • Testosterone (transmales) Doses and formulations not reported	**Critical Outcomes** *Impact on mental health* • Depression- The Center for Epidemiologic Studies Depression Scale (CESD-R) • Depression- The Patient Health Questionnaire Modified for Teens (PHQ 9_Modified for Teens) *Impact on quality of life* • Quality of Life Enjoyment and Satisfaction Questionnaire (QLES-Q-SF) **Important Outcomes** *None reported*

Study	Population	Intervention and comparison	Outcomes reported
		After about 12-months treatment ('wave 3'): • 24 people (48%) were on gender-affirming hormones alone • 12 people (24%) were on puberty suppression alone • 11 people (22%) were on both gender-affirming hormones and puberty suppression • 3 people (6%) were on no endocrine intervention **Comparison** No comparison group. Change over time reported	
Allen et al. 2019 Retrospective longitudinal study Single centre, Kansas City, USA	47 adolescents and young adults with gender dysphoria: 14 transfemales and 33 transmales Mean age at administration (start of treatment) 16.5 years	**Intervention** 39 participants received gender-affirming hormones only 8 participants received hormones and a GnRH analogue Mean duration of treatment with gender-affirming hormones was 349 days (range 113 to 1,016) **Comparison** No comparison group. Comparison over time reported	**Critical Outcomes** *Impact on mental health* • Suicidality- Ask Suicide-Screening Questions (ASQ) instrument *Impact on quality of life* • General Well-Being Scale (GWBS) of the Pediatric Quality of Life Inventory **Important Outcomes** *None reported*
Kaltiala et al. 2020 Retrospective chart review Single centre, Tampere, Finland	52 adolescents with gender dysphoria: 11 transfemales and 41 transmales. Mean age at diagnosis 18.1 years (range 15.2 to 19.9)	**Intervention** Hormonal sex assignment treatment – details of intervention not reported, although all patients received gender-affirming hormones.	**Critical Outcomes** *Impact on mental health* • Need for mental health treatment **Important Outcomes** *Psychosocial Impact* Measure of functioning in different domains of

Study	Population	Intervention and comparison	Outcomes reported
		Comparison No comparison group. Comparison over time reported	adolescent development, which were: • Living with parent(s)/ guardians • Normative peer contacts • Progresses normatively in school/ work • Has been dating or had steady relationships • Is age-appropriately able to deal with matters outside of the home
Khatchadourian et al. 2014 Retrospective chart review Single centre, Vancouver, Canada	84 young people with gender dysphoria, of whom 63 received gender-affirming hormones. Median age at start of gender-affirming hormones was: • 17.3 years (range 13.7-19.8) for testosterone • 17.9 years (range 13.3-22.3) for oestrogen	**Intervention** Transfemales: Oestrogen (oral micronized 17β-oestradiol) Transmales: Testosterone (injectable testosterone enanthate and/or cypionate) 19 participants (30%) had previously received a GnRH analogue **Comparison** No comparison group. Comparison over time reported.	**Critical Outcomes** *None reported* **Important Outcomes** *Safety:* • Adverse events • Discontinuation rates
Klaver et al. 2020 Retrospective chart review Single centre, Amsterdam, Netherlands	192 people with gender dysphoria who started GnRH analogues before the age of 18 years, and started gender-affirming hormones within 1.5 years of their 22nd birthday. Mean age at start of gender-affirming hormones: • Transfemale – 16.4 years (SD 1.1) • Transmale – 16.9 years (SD 1.9)	**Intervention** Oral oestrogen or intramuscular (IM) testosterone **Comparison** No comparison group. Comparison over time reported	**Critical Outcomes** *None reported* **Important Outcomes** *Safety* • Body mass index (BMI) • Systolic blood pressure • Diastolic blood pressure • Glucose • Insulin • HOMA-IR

Study	Population	Intervention and comparison	Outcomes reported
			• Total cholesterol • HDL cholesterol • LDL cholesterol • Triglycerides
Klink et al. 2015 Retrospective longitudinal study Single centre, Amsterdam, Netherlands	34 young people with gender dysphoria who had received GnRH analogues, gender-affirming hormones and gonadectomy. The study included 15 transfemales and 19 transmales; mean age at start of gender-affirming hormones was 16.6 years (SD 1.4) and 16.4 years (SD 2.3) respectively. At the start of gender-affirming hormone treatment, in the transfemale subgroup the median Tanner P was 4 (IQR 2) and the median Tanner G was 12 (IQR 11) In the transmale subgroup the median Tanner B was 5 (IQR 2) and the median Tanner P was 5 (IQR 0)	**Intervention** Transfemales – oral 17-β oestradiol (incremental dosing) Transmales – IM testosterone (Sustanon 250 mg/ml; incremental dosing) Median duration of treatment with gender-affirming hormones for transfemales was 5.8 years (range 3.0 to 8.0) and for transmales was 5.4 years (range 2.8 to 7.8) The GnRH analogue was subcutaneous (SC) triptorelin 3.75 mg every 4 weeks No details of gonadectomy reported **Comparison** No comparison group. Comparison over time reported.	**Critical Outcomes** None **Important Outcomes** *Safety* • Bone mineral apparent density (BMAD) • Bone mineral density (BMD) Measures reported at 3 timepoints: start of GnRH analogue treatment, start of gender-affirming hormone treatment and age 22 years.
Kuper et al. 2020 Prospective longitudinal study Single centre, Texas, USA	Children and adolescents with gender dysphoria (9 to18 years), n=148, of whom: • 25 received puberty suppression only • 93 received gender-affirming hormone therapy only • 30 received both Mean age 14.9 years	**Intervention** Gender-affirming hormones, guided by Endocrine Society Clinical Practice Guidelines **Comparison** No comparison group. Comparison over time reported.	**Critical Outcomes** *Impact on mental health* • Depression- Quick Inventory of Depressive Symptoms (QIDS), self-reported • Depression- QIDS, clinician-reported • Anxiety- Screen for Child Anxiety Related Emotional Disorders (SCARED)

Study	Population	Intervention and comparison	Outcomes reported
			• Panic- specific questions from SCARED • Generalised anxiety-specific questions from SCARED • Social anxiety - specific questions from SCARED • Separation anxiety-specific questions from SCARED • School avoidance-specific questions from SCARED **Important Outcomes** *Impact on body image* • Body Image Scale (BIS)
Lopez de Lara et al. 2020 Prospective analytical study Single centre, Madrid, Spain	23 adolescents with gender dysphoria: 7 transfemales and 16 transmales. Mean age at baseline was 16 years (range 14 to 18)	**Intervention** Gender-affirming hormones: • Oral oestradiol • Intramuscular testosterone Participants had previously received GnRH analogues in the intermediate pubertal stages (Tanner 2 to 3). Participants were assessed twice: • pre-treatment (T0), • after 12 months treatment with gender-affirming hormones (T1) **Comparison** No comparison group. Comparison over time reported.	**Critical Outcomes** *Impact on gender dysphoria* • Utrecht Gender Dysphoria Scale (UGDS) *Impact on mental health* • Depression- Beck Depression Inventory II (BDI-II) • Anxiety- State-Trait Anxiety Inventory **Important Outcomes** *Psychosocial Impact* • Family functioning-Family APGAR test • Patient strengths and difficulties-Strengths and Difficulties Questionnaire, Spanish Version (SDQ-Cas).
Stoffers et al. 2019 Retrospective chart review	62 transmales with gender dysphoria. Patients had received a GnRH analogue and more than 6 months of testosterone treatment.	**Intervention** Testosterone intramuscular injections (Sustanon 250 mg). Dose was titrated to a	**Critical Outcomes** None **Important Outcomes** *Safety*

Study	Population	Intervention and comparison	Outcomes reported
Single centre, Leiden, Netherlands	Median age at start of testosterone was 17.23 years (range 14.9 to 18.4) Median treatment duration was 12 months (range 5 to 33) Change over time	maintenance dose of 125 mg every 2 weeks. Participants who started GnRH analogues at 16 years or older had their dose increased more rapidly. Some participants chose to receive testosterone every 3-4 weeks, and participants could switch to transdermal preparations if needed. **Comparison** No comparison group. Comparison over time reported.	• Body mass index (BMI) • Blood pressure • BMD • Acne • Liver enzymes • Creatinine • Urea • HbA1c
Vlot et al. 2017 Retrospective chart review Single centre, Amsterdam, Netherlands	70 children and adolescents with gender dysphoria Median age at baseline – • 13.5 years (11.5-18.3) for transfemales • 15.1 years (range 11.7-18.6) for transmales Comparison is change over time. 24 month follow-up.	**Intervention** Oestrogen or testosterone (had previously received triptorelin for puberty suppression) **Comparison** No comparison group. Comparison over time reported.	**Critical Outcomes** None **Important Outcomes** *Safety* • Bone mineral apparent density (BMAD)

5. Results

In children and adolescents with gender dysphoria, what is the clinical effectiveness of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement
Clinical Effectiveness	
Critical outcomes	
Impact on gender dysphoria **Certainty of evidence: very low**	This is a critical outcome because gender dysphoria in children and adolescents is associated with significant distress and problems with functioning. One uncontrolled, prospective, observational study (Lopez de Lara et al. 2020) provided evidence relating to the impact on gender dysphoria, measured using the Utrecht Gender Dysphoria Scale (UGDS) score

	during the first year of treatment with gender-affirming hormones. The UGDS is a validated, screening tool for both adolescents and adults, used to assess gender dysphoria. It consists of 12 items, to be answered on a 1- to 5-point scale, resulting in a sum score between 12 and 60. The authors state that the cut-off point to identify gender dysphoria is 40 points. The higher the UGDS score the greater the gender dysphoria. In this study (n=23), the mean (±SD) UGDS score was statistically significantly reduced (improved) from 57.1 (±4.1) points at baseline to 14.7 points (±3.2) at 12 months (p<0.001). A UGDS score below 40 suggests an absence of gender dysphoria **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly improve gender dysphoria from baseline to 12 months follow-up. The mean UGDS score was below the threshold for gender dysphoria at follow-up.**
Impact on mental health: depression **Certainty of evidence: very low**	This is a critical outcome because depression may impact on social, occupational, or other areas of functioning in children and adolescents. Four observational studies (Achille et al. 2020; Kaltiala et al. 2020; Kuper et al. 2020; Lopez de Lara et al. 2020) provided evidence relating to the impact on depression in children and adolescents with gender dysphoria, with follow-up of around 12 months. Five different outcome measures for depression were reported. **Beck Depression Inventory (BDI-II)** One uncontrolled, prospective, analytical study (Lopez de Lara et al. 2020) reported the change in BDI-II. The BDI-II is a valid, reliable, and widely used tool for assessing depressive symptoms. There are no specific scores to categorise depression severity, but it is suggested that 0 to 13 is minimal symptoms, 14 to 19 is mild depression, 20 to 28 is moderate depression, and severe depression is 29 to 63. In Lopez de Lara et al. 2020 (n=23) the mean (±SD) BDI-II score was statistically significantly reduced (improved) from 19.3 (±5.5) points at baseline to 9.7 (±3.9) points at 12 months (p<0.001) **(VERY LOW)**. **Center for Epidemiologic Studies Depression (CESD-R)** One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in CESD-R scale. The CESD-R is a valid, widely used tool to assess depressive symptoms. Total score ranges from 0 to 60, with higher scores indicating more depressive symptoms. There are no specific scores to categorise depression severity, although the authors of the study suggest that a total CESD-R score less than 16 suggests no clinical depression. In Achille et al. 2020 (n=50), the mean CESD-R score statistically significantly reduced (improved) from 21.4 points at baseline to 13.9 points at about 12 months follow-up (p<0.001; standard deviation not reported) **(VERY LOW)**. **Patient Health Questionnaire (PHQ 9) Modified for Teens** One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in PHQ 9_Modified for Teens score. The PHQ

	9_Modified for Teens is a validated tool to assess depression, dysthymia and suicide risk. The tool consists of 9 questions scored from 0 to 3 (total score 0 to 27), plus an additional 4 questions that are not scored. A score of 0 to 4 suggests no or minimal depressive symptoms, 5 to 9 mild, 10 to 14 moderate, 15 to 19 moderately severe, and 20-27 severe symptoms. In Achille et al. 2020 (n=50), the mean PHQ 9_Modified for Teens score statistically significantly reduced (improved) from baseline to around 12 months follow-up, although absolute scores were not reported numerically ($p<0.001$). From the visual representation of results, the PHQ-9_Modified for Teens score is about 9 at baseline and about 5 at final follow-up **(VERY LOW)**. **Quick Inventory of Depressive Symptoms (QIDS)** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported the change in QIDS, clinician-reported and self-reported. Both the clinician-reported and self-reported QIDS are validated tools to assess depressive symptoms. The tool consists of 16 items, with the highest score for 9 domains (sleep, weight, psychomotor changes, depressed mood, decreased interest, fatigue, guilt, concentration, and suicidal ideation) added to give a total score ranging from 0 to 27. A score of 0 to 5 suggests no depression, 6 to 10 mild symptoms, 11 to 15 moderate symptoms, 16 to 20 severe symptoms, and 21 to 27 very severe symptoms. In Kuper et al. 2020 (n=105), the mean (±SD) QIDS self-reported score was 9.6 points (±5.0) at baseline and 7.4 (±4.5) after 10.9 months of treatment with gender-affirming hormones (no statistical analysis reported). The mean (±SD) QIDS clinician-reported score was 5.9 points (±4.1) at baseline and 6.0 (±3.8) after 10.9 months of treatment with gender-affirming hormones (no statistical analysis was reported) **(VERY LOW)**. **Participants needing treatment for depression** One observational study (Kaltiala et al. 2020) reported the proportion of participants needing treatment for depression before or during the initial assessment and during the 12-month follow-up period after starting gender-affirming hormones. In Kaltiala et al. 2020 (n=52), statistically significantly fewer participants needed treatment for depression during the 12-month 'real life' phase (15%, 8/52) compared with before or during the assessment (54%, 28/52; $p<0.001$). No details of what treatments for depression the participants received are reported **(VERY LOW)**. **These studies provide very low certainty evidence that during treatment with gender-affirming hormones depression is reduced from baseline to about 12 months follow-up. However, most participants had mild symptoms at the start of treatment.**
Impact on mental health: anxiety	This is a critical outcome because anxiety may impact on social, occupational, or other areas of functioning in children and adolescents.

Certainty of evidence: very low	Three observational studies (Kaltiala et al. 2020; Kuper et al. 2020; Lopez de Lara et al. 2020) provided evidence relating to the impact on anxiety in children and adolescents with gender dysphoria. **State-Trait Anxiety Inventory (STAI)** One uncontrolled, prospective, analytical study (Lopez de Lara et al. 2020) reported the change in STAI scores. STAI is a validated and commonly used measure of trait and state anxiety. It has 20 items and can be used in clinical settings to diagnose anxiety and to distinguish it from depressive illness. Higher scores indicate greater anxiety. In Lopez de Lara et al. 2020 (n=23), the mean (±SD) STAI-State subscale was statistically significantly reduced (improved) with gender-affirming hormones from 33.3 points (±9.1) at baseline to 16.8 points (±8.1) at 12 months ($p<0.001$). The mean STAI-Trait subscale scores also statistically significantly reduced (improved) from 33.0 points (±7.2) at baseline to 18.5 points (±8.4) at 12 months ($p<0.001$) **(VERY LOW)**. **Screen for Child Anxiety Related Emotional Disorders (SCARED)** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported anxiety symptoms using the SCARED questionnaire. Other anxiety-related symptoms using specific questions from the SCARED questionnaire were also reported: panic, generalised anxiety, social anxiety, separation anxiety and school avoidance. SCARED is a validated, 41-point questionnaire, with each item scored 0 to 2. A total score of 25 or more is suggestive of anxiety disorder, with scores above 30 being more specific. Certain scores for specific questions may indicate the presence of other anxiety-related disorders: • A score of 7 or more in questions related to panic disorder or significant somatic symptoms may indicate the presence of these. • A score of 9 or more in questions related to generalised anxiety disorder may indicate the presence of this. • A score of 5 or more in questions related to separation anxiety may indicate the presence of this. • A score of 8 or more in questions related to social anxiety disorder may indicate the presence of this. • A score of 3 or more in questions related to significant school avoidance may indicate the presence of this. In Kuper et al. 2020 (n=80 to 82, varies by outcome), small reductions were seen in anxiety, panic, generalised anxiety, social anxiety and separation anxiety and school avoidance symptoms (measured using the SCARED questionnaire) from baseline to follow-up (mean duration of treatment 10.9 months). The statistical significance of these findings are unknown as no statistical analyses were reported **(VERY LOW)**. **Participants needing treatment for anxiety** One observational study (Kaltiala et al. 2020) reported the proportion of participants needing treatment for anxiety before or during initial assessment and during the 12-month follow-up period after starting gender-affirming hormones.

	In Kaltiala et al. 2020 (n=52), statistically significantly fewer participants needed treatment for anxiety during the 12-month 'real life' phase (15%, 8/52) compared with before or during the assessment (48%, 25/52; p<0.001). No details of what treatments for anxiety the participants received are reported **(VERY LOW)**. **These studies provide very low certainty evidence that during treatment with gender-affirming hormones anxiety symptoms may be reduced from baseline to around 12 months follow-up.**
Impact on mental health: suicidality and self-injury **Certainty of evidence: very low**	These are critical outcomes because self-harm and thoughts of suicide have the potential to result in significant physical harm and, for completed suicides, the death of the young person. Four observational studies (Achille et al. 2020; Allen et al. 2019; Kaltiala et al. 2020; Kuper et al. 2020) provided evidence relating to suicidal ideation in children and adolescents with gender dysphoria, with an average follow-up of around 12 months. **Ask Suicide-Screening Questions (ASQ)** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in ASQ. This is a 4-item dichotomous (yes/no) response measure designed to identify risk of suicide. The authors of Allen et al. 2019 amended 1 question in the ASQ ("*Have you ever tried to kill yourself?*") by prefacing it with "*In the past few weeks . . .*" as they were not investigating lifetime incidence. A response of 'no' is scored as 0 and a response of 'yes' is scored as 1; each item is summed to give an overall score for suicidal ideation ranging from 0 to 4. A person is considered to have screened positive if they answer 'yes' to any item with higher scores indicating higher levels of suicidal ideation. In Allen et al. 2019 (n=39), the adjusted mean (±SE) ASQ score statistically significantly reduced from 1.11 points (±0.22) at baseline to 0.27 points (±0.12) after a mean duration of treatment of about 12 months (p<0.001) **(VERY LOW)**. **PHQ 9_Modified for Teens (additional questions for suicidal ideation)** One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in suicidal ideation measured using additional questions from the PHQ 9_Modified for Teens. This is a validated tool to assess depression, dysthymia and suicide risk (see above for detailed description). In addition to the 9 scored questions, the PHQ 9_Modified Teens asked 4 additional questions relating to suicidal ideation and difficulty dealing with problems of life. Responses to the PHQ 9_Modified for Teens were used to determine if the participant had suicidal ideation or not, but specific details of how this was determined are not reported. In Achille et al. 2020 (n=50), 10% (5/50) of participants had suicidal ideation at baseline and 6% (3/50) had suicidal ideation after about 12 months treatment with gender-affirming hormones (no statistical analysis reported) **(VERY LOW)**. **Suicidality and non-suicidal self-injury**

	One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported on suicidal ideation, suicide attempts and non-suicidal self-injury, although it was unclear how and when this outcome was measured. In Kuper et al. 2020 (n=130), 25% of participants reported suicidal ideation 1 month before the initial assessment and 38% reported this during the follow-up period (no statistical analysis reported). Suicide attempts were reported in 2% of participants at 3 months before the initial assessment and 5% during follow-up. Self-injury was reported in 10% of participants at 3 months before the initial assessment and 17% during follow-up. No statistical analysis was reported for any outcomes. Mean duration of gender-affirming hormone treatment was 10.9 months **(VERY LOW)**. **Participants needing treatment for suicidality or self-harm** One observational study (Kaltiala et al. 2020) reported the proportion of participants requiring treatment for suicidality or self-harm before or during initial assessment and during the 12-month follow-up period after starting gender-affirming hormones. In Kaltiala et al. 2020 (n=52) statistically significantly fewer participants needed treatment for suicidality or self-harm during the 12-month 'real life' phase (4%, 2/52) compared with before or during the assessment (35%, 18/52; p<0.001). No details of what treatments for suicidal ideation or self-harm the participants received are reported **(VERY LOW)**. **These studies provide very low certainty evidence that gender-affirming hormones may reduce suicidality from baseline to about 12 months follow-up. However, results are inconsistent and it is difficult to draw conclusions.**
Impact on mental health: other **Certainty of evidence: very low**	This is a critical outcome because mental health problems may impact on social, occupational, or other areas of functioning in children and adolescents. One observational study (Kaltiala et al. 2020) reported the proportion of participants needing treatment for either psychotic symptoms or psychosis, substance abuse, autism, attention deficit hyperactivity disorder (ADHD) or eating disorders before or during initial assessment and during the 12-month follow-up period after starting gender-affirming hormones. In Kaltiala et al. 2020 (n=52) there was no statistically significant difference in the number of people needing treatment for either psychotic symptoms / psychosis, substance abuse, autism, attention deficit hyperactivity disorder (ADHD) or eating disorders during the 12-month 'real life' phase compared with before or during the assessment. No details of which specific treatments the participants received are reported **(VERY LOW)**. **This study provides very low certainty evidence on the need for treatment for either psychotic symptoms or psychosis, conduct problems or antisocial behaviour, substance abuse, autism, attention deficit hyperactivity disorder (ADHD) or eating disorders**

	during treatment with gender-affirming hormones. No conclusions could be drawn.
Impact on quality of life score **Certainty of evidence: very low**	This is a critical outcome because gender dysphoria in children and adolescents may be associated with a significant reduction in health-related quality of life. Two uncontrolled longitudinal studies Achille et al. 2020; Allen et al. 2019) provided evidence relating to quality of life in children and adolescents with gender dysphoria. **Quality of Life Enjoyment and Satisfaction Questionnaire (QLES-Q-SF)** One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in QLES-Q-SF scores from baseline to about 12 months of treatment with gender-affirming hormones. QLES-Q-SF is a validated questionnaire, consisting of 15 questions that rate quality of life on a scale of 1 (poor) to 5 (very good). In Achille et al. 2020 (n=50), the mean QLES-Q-SF score was statistically significantly reduced from baseline to about 12 months ($p<0.001$). However, absolute scores are not reported numerically **(VERY LOW)**. **General Well-Being Scale (GWBS) of the Paediatric Quality of Life Inventory** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in adjusted mean GWBS of the Paediatric Quality of Life Inventory score from baseline to about 12 months of treatment with gender-affirming hormones. The GWBS of the Paediatric Quality of Life Inventory contains 7 items that measure two dimensions: general wellbeing (6 items) and general health (1 item). Each item is scored from 0 to 4, and the total score is linearly transformed to a 0 to 100 scale. Higher scores reflect fewer perceived problems and greater well-being. In Allen et al. 2019 (n=47), the adjusted mean (±SE) GWBS of the Paediatric Quality of Life Inventory score was statistically significantly increased (improved) from 61.70 (±2.43) points at baseline to 70.23 (±2.15) points at about 12 months ($p<0.002$) **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly improve quality of life and well-being from baseline to 12 months follow-up.**
Important outcomes	
Impact on body image **Certainty of evidence: very low**	This is an important outcome because some children and adolescents with gender dysphoria may want to take steps to suppress features of their physical appearance associated with their sex assigned at birth or accentuate physical features of their desired gender. One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) provided evidence relating to the impact on body image in children and adolescents with gender dysphoria who started treatment with gender-affirming hormones (median duration 10.9 months; range 1 to 18), measured by the change in Body Image Scale (BIS) score. BIS is a

	validated 30-item scale covering 3 aspects: primary, secondary and neutral body characteristics. Higher scores represent a higher degree of body dissatisfaction. In Kuper et al. 2020 (n=86), the mean (±SD) BIS score was 70.7 points (±15.2) at baseline and 51.4 points (±18.3) at follow-up (no statistical analysis reported) **(VERY LOW)**. **This study provides very low certainty evidence on the effects of gender-affirming hormones on body image during treatment with gender-affirming hormones (mean duration of treatment 10.9 months). No conclusions could be drawn.**
Psychosocial impact **Certainty of evidence: very low**	This is an important outcome because gender dysphoria in children and adolescents is associated with internalising and externalising behaviours, and emotional and behavioural problems which may impact on social and occupational functioning. Two uncontrolled, observational studies (Kaltiala et al. 2020; Lopez de Lara et al. 2020) provided evidence related to psychosocial impact in children and adolescents with gender dysphoria. **Family APGAR (Adaptability, Partnership, Growth, Affection and Resolve) test** One uncontrolled, prospective, analytical study (Lopez de Lara et al. 2020) reported the Family APGAR test. The Family APGAR test is a 5-item questionnaire, with higher scores indicating better family functioning. The authors reported the following interpretation of the test: functional, 17 to 20 points; mildly dysfunctional, 16 to 13 points; moderately dysfunctional, 12 to 10 points; severely dysfunctional, <9 points. In Lopez de Lara et al. 2020 (n=23), the mean Family APGAR test score was unchanged from baseline (17.9 points) to 12-month follow-up (18.0 points; no statistical analysis or standard deviations reported) **(VERY LOW)**. **Strengths and Difficulties Questionnaire (SDQ)** One uncontrolled, prospective, analytical study (Lopez de Lara et al. 2020) reported on behaviour using the Strengths and Difficulties Questionnaire (SDQ, Spanish version). The SDQ includes 25-items covering emotional symptoms, conduct problems, hyperactivity/ inattention, peer relationship problems and prosocial behaviour. The authors state that a score of more than 20 suggests having a behavioural disorder (normal 0 to 15, borderline 16 to 19, abnormal 20 to 40). In Lopez de Lara et al. 2020 (n=23), the mean (±SD) SDQ score was statistically significantly reduced (improved) from 14.7 points (±3.3) at baseline to 10.3 points (±2.9) at 12-month follow-up ($p<0.001$) **(VERY LOW)**. **Psychosocial functioning** One uncontrolled, retrospective chart review (Kaltiala et al. 2020) reported various markers of functioning in adolescent development, covering living arrangements, peer contacts, school or work progress,

	relationships, and ability to cope with matters outside the home. These measures were reported during the gender identity assessment and at about 12 months after starting gender-affirming hormones (referred to as the 'real-life phase'). In Kaltiala et al. 2020 (n=52), from the gender identity assessment to the 12-month follow-up period: • statistically significantly fewer participants were living with parents or guardians (73% versus 40%, p=0.001) • statistically significantly fewer participants had normal peer contacts (89% versus 81%, p<0.001) • there was no statistically significant difference in progress in school or work (64% versus 60%, p=0.69) • there was no statistically significant difference in the number of participants who had been dating or in steady relationships (62% versus 58%, p=0.51) • there was no statistically significant difference in the participant's ability to cope with matters outside of the home (81% versus 81%, p=1.00) **(VERY LOW)**. **These studies provide very low certainty evidence that gender-affirming hormones statistically significantly improve behavioural problems (measured by SDQ score). However, the SDQ score was in the 'normal' range at baseline and at 12-month follow up. There was no significant impact on other measures of psychosocial functioning.**
Engagement with health care services	This is an important outcome because patient engagement with health care services will impact on their clinical outcomes. No evidence was identified.
Impact on extent of and satisfaction with surgery	This is an important outcome because some children and adolescents with gender dysphoria may proceed to transitioning surgery. No evidence was identified.
De-transition	This is an important outcome because there is uncertainty about the short- and long-term safety and adverse effects of gender-affirming hormones in children and adolescents with gender dysphoria No evidence was identified.

Abbreviations: APGAR: Adaptability, Partnership, Growth, Affection and Resolve; ASQ: Ask Suicide-Screening Questions; BDI-II: Beck Depression Inventory II; BIS: Body Image Scale; CESD-R: Center for Epidemiologic Studies Depression; GWBS: General Well-Being Scale; p: p-value; PHQ 9_Modified for Teens: Patient Health Questionnaire Modified for Teens; QIDS: Quick Inventory of Depressive Symptoms; QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire; SCARED: Screen for Child Anxiety Related Emotional Disorders; SD: standard deviation; SE: standard error; SDQ: Strengths and Difficulties Questionnaire; STAI: State-Trait Anxiety Inventory; UGDS: Utrecht Gender Dysphoria Scale.

In children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement
Safety	
Change in bone density: lumbar spine **Certainty of evidence: very low**	This is an important outcome because childhood and adolescence is a key time for bone development and gender-affirming hormones may affect bone development, as shown by changes in lumbar spine bone density. Three uncontrolled, observational studies (2 retrospective and 1 prospective) provided evidence related to bone density: lumbar spine in children and adolescents with gender dysphoria. This was reported as either bone mineral density (BMD), bone mineral apparent density (BMAD), or both. One study reported change in bone density from start of treatment with gender-affirming hormones to age 22 years (Klink et al. 2015). Two studies reported change in bone density from start of gender-affirming hormones up to 24-month follow-up (Stoffers et al. 2019 and Vlot et al. 2017). All participants had previously been treated with a GnRH analogue. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. **Bone mineral apparent density (BMAD)** Two uncontrolled, observational studies reported change in lumbar BMAD (Klink et al. 2015; Vlot et al. 2017). BMAD is a size adjusted value of BMD, incorporating bone size measurements using a UK reference population of growing cis-gender adolescents (up to age 17 years). BMAD is used to correct for height and height gain and may provide a more accurate estimate of bone density in growing adolescents. BMAD was reported as g/cm^3 and as z-scores. Z-scores report how many standard deviations from the mean a measurement sits. A z-score of 0 is equal to the mean, a z-score of -1 is equal to 1 standard deviation below the mean, and a z-score of +1 is equal to 1 standard deviation above the mean. A cis-gender population was used to calculate the bone density z-score, meaning transfemales were compared with cis-males and transmales were compared with cis-females. In Klink et al. 2015 (n=34): • There was no statistically significant difference in lumbar spine BMAD z-score from starting gender-affirming hormones to age 22 years in transfemales. • The z-score for lumbar spine BMAD was statistically significantly higher at age 22 years compared with the start of gender-affirming hormones in transmales (z-score [±SD]: start of hormones -0.50 [±0.81], age 22 years -0.033 [±0.95], p=0.002). • Actual lumbar spine BMAD values in g/cm^3 were statistically significantly higher at age 22 years compared with the start of gender-affirming hormones in transfemales and transmales **(VERY LOW)**. In Vlot et al. 2017 (n=70): • The z-score for lumbar spine BMAD in transfemales with a bone age of <15 years was statistically significantly higher at 24-month follow-up compared with start of gender-affirming hormones (z-score [range]: start of hormones -1.52 [-2.36 to

	0.42], 24-month follow-up -1.10 [-2.44 to 0.69], p≤ 0.05). Statistically significant improvements in z-score for lumbar spine BMAD in transfemales with a bone age of ≥15 years were also seen (z-score [range]: start of hormones -1.15 [-2.21 to 0.08], 24-month follow-up -0.66 [-1.66 to 0.54], p≤ 0.05). • The z-score for lumbar spine BMAD in transmales with a bone age of <14 years was statistically significantly higher at 24-month follow-up compared with start of gender-affirming hormones (z-score [range]: start of hormones -0.84 [-2.2 to 0.87], 24-month follow-up -0.15 [-1.38 to 0.94], p≤ 0.01). Statistically significant improvements in z-score for lumbar spine BMAD in transmales with a bone age of ≥14 years were also seen (z-score [range]: start of hormones -0.29 [-2.28 to 0.90], 24-month follow-up -0.06 [-1.75 to 1.61], p≤ 0.01). • Actual lumbar spine BMAD values in g/cm^3 were statistically significantly higher at 24-month follow-up compared with start of gender-affirming hormones in transfemales and transmales of all bone ages **(VERY LOW)**. **Bone mineral density (BMD)** Two uncontrolled, observational studies reported change in lumbar BMD (Klink et al. 2015; Stoffers et al. 2019). BMD was determined using dual energy x-ray absorptiometry (DXA-scan; HologicQDR4500, Hologic). BMD was reported as g/cm^2 and as z-scores – see BMAD above for more details). In Klink et al. 2015 (n=34): • There was no statistically significant difference in lumbar spine BMD z-score from starting gender-affirming hormones to age 22 years in transfemales or transmales. • Actual lumbar spine BMD values in g/cm^2 were statistically significantly higher at age 22 years compared with the start of gender-affirming hormones in transfemales and transmales **(VERY LOW)**. In Stoffers et al. 2019 (n=62 at 6-month follow-up; n=15 at 24-month follow-up): • There was no statistically significant difference in lumbar spine BMD z-score in transmales from starting gender-affirming hormones to any timepoint (6, 12 and 24 months). • There was also no statistically significant difference in actual lumbar spine BMD values in g/cm^2 from starting gender-affirming hormones to any timepoint (6, 12 and 24 months) **(VERY LOW)**. **These studies provide very low certainty evidence that lumber spine bone density (measured by BMAD) increases during treatment with gender-affirming hormones (from baseline to follow-up of 2 to 5 years). Z-scores at the end of follow-up suggest the average lumbar spine bone density was generally lower than the equivalent cisgender population (transfemales compared with cis-males and transmales compared with cis-females). The results for bone density (measured by BMD) were inconsistent.**

Change in bone density: femoral neck **Certainty of evidence: very low**	This is an important outcome because childhood and adolescence is a key time for bone development and gender-affirming hormones may affect bone development, as shown by changes in femoral neck bone density. Three uncontrolled, observational studies (2 retrospective and 1 prospective) provided evidence related to bone density: femoral neck in children and adolescents with gender dysphoria. This was reported as either bone mineral density (BMD), bone mineral apparent density (BMAD), or both. One study reported change in bone density from start of gender-affirming hormones to age 22 years (Klink et al. 2015). Two studies reported change in bone density from start of gender-affirming hormones up to 24-month follow-up (Stoffers et al. 2019 and Vlot et al. 2017). All participants had previously been treated with a GnRH analogue. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. **Bone mineral apparent density (BMAD)** Two uncontrolled, observational studies reported change in femoral neck BMAD (Klink et al. 2015; Vlot et al. 2017). See above for more details on BMAD. In Klink et al. 2015 (n=34): • The z-score for femoral neck BMAD was reported for the start of gender-affirming hormones but not at age 22 years in transfemales or transmales. No statistical analysis reported. • In transfemales there was no statistically significant difference in actual femoral neck BMAD values in g/cm^3 at age 22 years compared with start of gender-affirming hormones. In transmales actual lumbar spine BMAD values in g/cm^3 were statistically significantly higher at age 22 years compared with start of gender-affirming hormones (mean [±SD]: start of hormones 0.31 [±0.04], age 22 years 0.33 [±0.05], p=0.010) **(VERY LOW)**. In Vlot et al. 2017 (n=70): • In transfemales (all bone ages), there was no statistically significant difference in femoral neck BMAD z-score from start of gender-affirming hormones to 24-month follow-up. • The z-score for femoral neck BMAD in transmales with a bone age of <14 years was statistically significantly higher at 24-month follow-up compared with start of gender-affirming hormones (z-score [range]: start of hormones -0.37 [-2.28 to 0.47], 24-month follow-up -0.37 [-2.03 to 0.85], p≤0.01). Statistically significant improvements in z-score for lumbar spine BMAD in transmales with a bone age of ≥14 years were also seen (z-score [range]: start of hormones -0.27 [-1.91 to 1.29], 24-month follow-up 0.02 [-2.1 to 1.35], p≤0.05). • In transfemales of all bone ages, there was no statistically significant change in actual femoral neck BMAD values in g/cm^3 from start of gender-affirming hormones to 24-month follow-up. In transmales of all bone ages, actual femoral neck BMAD values in g/cm^3 were statistically significantly higher at 24-month follow-up compared with start of gender-affirming hormones **(VERY LOW)**.

	Bone mineral density (BMD) Two uncontrolled, observational studies reported change in femoral neck BMD (Klink et al. 2015; Stoffers et al. 2019). See above for more details on BMD. In Klink et al. 2015 (n=34): ▪ In transfemales, there was no statistically significant difference in femoral neck BMD z-score from start of gender-affirming hormones to age 22 years. In transmales, femoral neck BMD z-score was statistically significantly higher at age 22 years compared with start of gender-affirming hormones (z-score [SD]: start of hormones -0.35 [0.79], age 22 years -0.35 [0.74], p=0.006). ▪ Actual femoral neck BMD values in g/cm^2 were statistically significantly higher at age 22 years compared with start of gender-affirming hormones in transfemales and transmales **(VERY LOW)**. In Stoffers et al. 2019 (n=62 at 6-month follow-up; n=15 at 24-month follow-up): ▪ there was no statistically significant difference in right or left femoral neck BMD z-score in transmales, from the start of gender-affirming hormones to any timepoint (6, 12 and 24 months). ▪ There was also no statistically significant difference in transmales in right or left actual femoral neck BMD values in g/cm^2 from start of gender-affirming hormones to any timepoint (6, 12 and 24 months) **(VERY LOW)**. **These studies provide very low certainty evidence that during treatment with gender-affirming hormones from baseline to follow-up of 2 to 5 years, femoral neck bone density (measured by BMAD) was unchanged in transfemales but was statistically significantly increased in transmales (although the absolute change was small). Z-scores at the end of follow-up suggest that average femoral neck bone density was lower in both transfemales and transmales than in the equivalent cisgender population (transfemales compared with cis-males and transmales compared with cis-females). The results for bone density (measured by BMD) were inconsistent.**
Change in clinical parameters: glucose, insulin and HbA1c **Certainty of evidence: very low**	This is an important outcome because the effect of gender-affirming hormones on insulin sensitivity and cardiovascular risk in children and adolescents with gender dysphoria is unknown. Two uncontrolled, retrospective chart reviews (Klaver et al. 2020; Stoffers et al. 2019) provided evidence on glucose, insulin and HbA1c. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. **Glucose levels, insulin levels and insulin resistance** One retrospective chart review (Klaver et al. 2020) reported non-comparative evidence on the change in glucose levels, insulin levels and insulin resistance (measured using Homeostatic Model

	Assessment of Insulin Resistance [HOMA-IR]) between starting gender-affirming hormones and age 22 years. In Klaver et al. 2020 (n=192): • There was no statistically significant change in glucose levels, insulin levels and insulin resistance in transfemales. • There was no statistically significant change in glucose levels in transmales. • There was a statistically significant decrease in insulin levels in transmales (mean change [95% CI] -2.1 mU/L [-3.9 to -0.3], p<0.05; mean insulin level at 22 years [95% CI] 8.6 mU/L [6.9 to 10.2]). • There was a statistically significant decrease in insulin resistance in transmales (HOMA-IR; mean change [95% CI] -0.5 [-1.0 to -0.1], p<0.05; mean HOMA-IR at 22 years [95% CI] 1.8 [1.4 to 2.2]) **(VERY LOW)**. **HbA1c** One retrospective chart review (Stoffers et al. 2019; n=62) reported non-comparative evidence on the change in HbA1c in transmales between starting gender-affirming hormones and 24-month follow-up. There was no statistically significant change in HbA1c **(VERY LOW)**. **These studies provide very low certainty evidence that gender-affirming hormones do not affect HbA1c, glucose levels, insulin levels and insulin resistance.**
Change in clinical parameters: lipids **Certainty of evidence: very low**	This is an important outcome because the effect of gender-affirming hormones on lipid profiles and cardiovascular risk in children and adolescents with gender dysphoria is unknown. One retrospective chart review (Klaver et al. 2020) provided non-comparative evidence on the change in lipids (total cholesterol, HDL cholesterol, LDL cholesterol and triglycerides) between starting gender-affirming hormones and age 22 years. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. In Klaver et al. 2020 (n=192): • There was no statistically significant change in total cholesterol, HDL cholesterol and LDL cholesterol in transfemales. • There was a statistically significant decrease (improvement) in triglycerides in transfemales (mean change [95% CI] +0.2 mmol/L [0.0 to 0.5], p<0.05; mean triglyceride level at 22 years [95% CI] 1.1 mmol/L [0.9 to 1.4]). • There was a statistically significant increase in total cholesterol in transmales (mean change [95% CI] +0.4 mmol/L [0.2 to 0.6], p<0.001; mean total cholesterol at 22 years [95% CI] 4.6 mmol/L [4.3 to 4.8]). • There was a statistically significant decrease (worsening) in HDL cholesterol (mean change in transmales [95% CI] -0.3 mmol/L [-0.4 to -0.1], p<0.001; mean HDL cholesterol at 22 years [95% CI] 1.3 mmol/L [1.2 to 1.3]). • There was a statistically significant increase (worsening) in LDL cholesterol in transmales (mean change [95% CI]

	+0.4 mmol/L [0.2 to 0.6], p<0.001; mean LDL cholesterol at 22 years [95% CI] 2.6 mmol/L [2.4 to 2.8]). • There was a statistically significant increase (worsening) in triglycerides in transmales (mean change [95% CI] +0.5 mmol/L [0.3 to 0.7], p<0.001; mean triglyceride level at 22 years [95% CI] 1.3 mmol/L [1.1 to 1.5]) **(VERY LOW)** **This study provides very low certainty evidence that gender-affirming hormones do not affect lipid profiles in transfemales. In transmales, there was a small but statistically significant worsening in cholesterol levels from start of gender-affirming hormone treatment to age 22 years, but mean cholesterol and triglyceride levels were within the UK reference range at the end of treatment.**
Change in clinical parameters: blood pressure **Certainty of evidence: very low**	This is an important outcome because the effect of gender-affirming hormones on blood pressure and cardiovascular risk in children and adolescents with gender dysphoria is unknown. One retrospective chart review (Klaver et al. 2020) provided non-comparative evidence on the change in blood pressure between starting gender-affirming hormones and at age 22 years. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. In Klaver et al. 2020 (n=192): • There was no statistically significant change in systolic blood pressure (SBP) in transfemales. However, there was a statistically significant increase in diastolic blood pressure (DBP) in transfemales (mean change [95% CI] +6 mmHg [3 to 10], p<0.001; mean DBP at 22 years [95% CI] 75 [72 to 78]). • In transmales, there was a statistically significant increase in SBP (mean change [95% CI] +5 mmHg [1 to 9], p<0.05; mean SBP at 22 years [95% CI] 126 [122 to 130]), and DBP (mean change [95% CI] +6 mmHg [4 to 9], p<0.001; mean DBP at 22 years [95% CI] 74 [72 to 77]) **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase blood pressure from start of treatment to age 22 years, although the absolute increase was small.**
Change in clinical parameters: body mass index (BMI) **Certainty of evidence: very low**	This is an important outcome because the effect of gender-affirming hormones on weight gain and cardiovascular risk in children and adolescents with gender dysphoria is unknown. One retrospective chart review (Klaver et al. 2020) provided non-comparative evidence on the change in body mass index (BMI) between starting gender-affirming hormones and age 22 years. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. In Klaver et al. 2020 (n=192): • There was a statistically significant increase in BMI in transfemales from the start of gender-affirming hormones to age 22 years (mean change [95% CI] +1.9 [0.6 to 3.2], p<0.005; mean BMI at 22 years [95% CI] 23.2 [21.6 to 24.8]. At age 22

	years, 9.9% of transfemales were obese, compared with 3.0% in a reference population of cisgender men. • There was a statistically significant increase in BMI in transmales from the start of gender-affirming hormones to age 22 years (mean change [95% CI] +1.4 [0.8 to 2.0], p<0.005; mean BMI at 22 years [95% CI] 23.9 [23.0 to 24.7]). At age 22 years, 6.6% of transmales were obese, compared with 2.2% in a reference population of cisgender women **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase BMI from start of treatment to age 22 years, although most participants were within the healthy weight range.**
Change in clinical parameters: liver function **Certainty of evidence: very low**	This is an important outcome because if treatment-induced liver injury (raised liver enzymes are a marker of this) is suspected, gender-affirming hormones may need to be stopped. One retrospective chart review (Stoffers et al. 2019) provided non-comparative evidence on the change in liver enzymes in transmales between starting gender-affirming hormones and up to 24-months follow-up. In Stoffers et al. 2019 (n=62): • There was no statistically significant change in aspartate aminotransferase (AST), alanine aminotransferase (ALT) and gamma-glutamyltransferase (GCT) in transmales. • There was a statistically significant increase in alkaline phosphatase (ALP) levels from starting gender-affirming hormones to 6- and 12-months follow-up, although by 24-months the difference was not statistically significant (median [IQR]: start of hormones 102 [78 to 136], 6-month follow-up 115 [102 to 147] p<0.001, 12-month follow-up 112 [88 to 143] p<0.001) **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones do not affect liver function in transmales from baseline to 24 months follow-up.**
Change in clinical parameters: kidney function **Certainty of evidence: very low**	This is an important outcome because if renal damage (raised serum creatinine and urea are markers of this) is suspected, treatment with gender-affirming hormones may need to be stopped. One retrospective chart review (Stoffers et al. 2019) provided non-comparative evidence on the change in serum creatinine and serum urea levels in transmales between starting gender-affirming hormones and up to 24-months follow-up. In Stoffers et al. 2019 (n=62): • There was a statistically significant increase in creatinine levels in transmales at all timepoints up to 24 months (mean [SD]: start of hormones 62 umol/L [7], 6 months 70 umol/L [9] , 12 months 74 umol/L [10], 24 months 81 umol/L [10], p<0.001) • There was no statistically significant change in urea in transmales (follow-up duration not reported) **(VERY LOW)**.

	This study provides very low certainty evidence on the effects of gender-affirming hormones on kidney function in transmales from baseline to 24 months follow-up. A statistically significant increase in creatinine levels was seen, but these were within the UK reference range. Urea levels were unchanged.

Treatment discontinuation
 	
Certainty of evidence: very low
	This is an important outcome because there is uncertainty about the short- and long-term impact of stopping treatment with gender-affirming hormones in children and adolescents with gender dysphoria.
	
	One uncontrolled, retrospective chart review (Khatchadourian et al. 2014) provided evidence relating to permanent or temporary treatment discontinuation in children and adolescents with gender dysphoria.
	
	Khatchadourian et al. 2014 narratively reported treatment discontinuation in a cohort of 63 adolescents (24 transfemales and 39 transmales) who received gender-affirming hormones:
	• No participants permanently discontinued gender-affirming hormones.
	• No transfemales temporarily discontinued gender-affirming hormones.
	• Three transmales temporarily discontinued gender-affirming hormones due to:
	  ○ mental health comorbidities (n=2)
	  ○ androgenic alopecia (n=1).
	All 3 participants eventually resumed treatment, although timescales were not reported **(VERY LOW)**.
	
	This study provides very low certainty evidence that the rates of discontinuation during treatment with gender-affirming hormones are low (duration of treatment not reported).

Adverse effects
 	
Certainty of evidence: very low
	This is an important outcome because if there are adverse effects, gender-affirming hormones may need to be stopped.
	
	One uncontrolled, retrospective chart review (Khatchadourian et al. 2014) provided evidence relating to adverse effects from gender-affirming hormones in children and adolescents with gender dysphoria.
	
	Khatchadourian et al. 2014 narratively reported adverse effects in a cohort of 63 adolescents (24 transfemales and 39 transmales) receiving treatment with gender-affirming hormones:
	• No severe complications were reported.
	• No transfemales reported minor complications.
	• Twelve transmales developed minor complications, which were:
	  ○ severe acne, requiring isotretinoin treatment (n=7)
	  ○ androgenic alopecia (n=1)
	  ○ mild dyslipidaemia (further details not provided; n=3)
	  ○ significant mood swings (n=1) **(VERY LOW)**.
	
	This study provides very low certainty evidence about the potential adverse effects of gender-affirming hormones (duration of treatment not reported). No conclusions could be drawn.

Abbreviations: ALP: alkaline phosphatase; ALT: alanine aminotransferase; AST: aspartate aminotransferase; BMAD: bone mineral apparent density; BMD: bone mineral density; BMI: body mass index; DBP: diastolic blood pressure; GGT: gamma-glutamyl transferase; HbA1c:

glycated haemoglobin; HDL: high-density lipoproteins; HOMA-IR: Homeostatic Model Assessment of Insulin Resistance; IQR: interquartile range; LDL: low-density lipoproteins; p: p-value; SBP: systolic blood pressure; SD: standard deviation.

In children and adolescents with gender dysphoria, what is the cost-effectiveness of gender-affirming hormones compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement
Cost-effectiveness	No studies were identified to assess the cost-effectiveness of gender-affirming hormones for children and adolescents with gender dysphoria.

From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may benefit from gender-affirming hormones more than the wider population of interest?

Subgroup	Evidence statement
Sex assigned at birth males (transfemales) **Certainty of evidence: Very low**	Some studies reported data separately for sex assigned at birth males (transfemales). This included some direct comparisons with sex assigned at birth females (transmales). **Impact on mental health: depression and anxiety** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported the change in depression (measured using QIDS clinician-reported and self-reported), anxiety and anxiety-related symptoms (measured using SCARED) in transfemales. See the clinical effectiveness results above for full details. In Kuper et al. 2020 (n=33 to 45, varies by outcome), changes were seen in depression, anxiety and anxiety-related symptoms from baseline to follow-up but the authors did not report any statistical analyses, so it is unclear if was any changes were statistically significant **(VERY LOW)**. **This study provides very low certainty evidence on the effects of gender-affirming hormones on depression, anxiety and anxiety-related symptoms over time in sex assigned at birth males (transfemales; mean duration of treatment 10.9 months). No conclusions could be drawn.** **Impact on mental health: suicidality** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in Ask Suicide-Screening Questions (ASQ) in transfemales compared with transmales. See the clinical effectiveness results above for full details. Between baseline and the final assessment, there was no statistically significant difference in change in ASQ score for transfemales compared with transmales (p=0.79; n=47) **(VERY LOW)**.

	One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in suicidal ideation in transfemales measured using additional questions from the PHQ 9_Modified for Teens. See the clinical effectiveness results above for full details. At baseline, 11.8% (2/17) of transfemales had suicidal ideation, compared with 5.9% (1/17) at about 12-months follow-up (no statistical analysis reported) **(VERY LOW)**. **These studies provide very low certainty evidence that any change in suicidal ideation is not different between sex assigned at birth males (transfemales) and sex assigned at birth females (transmales) from baseline to follow-up of about 12 months.** **Impact on quality of life** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in the GWBS of the Paediatric Quality of Life Inventory in transfemales compared with transmales. See the clinical effectiveness results above for full details. Between baseline and final assessment, there was no statistically significant difference in change in GWBS of the Paediatric Quality of Life Inventory for transfemales compared with transmales (p=0.32; n=47) **(VERY LOW)**. **This study provides very low certainty evidence that any change in general wellbeing is not different between sex assigned at birth males (transfemales) and sex assigned at birth females (transmales) from baseline to follow-up of about 12 months.** **Impact on body image** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported change in Body Image Scale (BIS) in transfemales. See the clinical effectiveness results above for full details. In Kuper et al. 2020 (n=30), the mean (±SD) BIS score was 67.5 points (±19.5) at baseline and 49.0 points (±21.6) at follow-up (no statistical analysis reported) **(VERY LOW)**. **This study provides very low certainty evidence on the effects of gender-affirming hormones on body image over time in transfemales (mean duration of treatment 10.9 months). No conclusions could be drawn.** **Change in bone density: lumbar spine** Two uncontrolled, observational, retrospective studies provided evidence relating to the effect of gender-affirming hormones on lumber spine bone density in transfemales (Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results.

	These studies provide very low certainty evidence that lumbar spine bone density (measured by BMAD) increases during treatment with gender-affirming hormones in sex assigned at birth males (transfemales). Z-scores at the end of follow-up suggest average lumbar spine bone density was generally lower than in the equivalent cisgender population. The results for lumbar spine bone density (measured by BMD) were inconsistent. **Change in bone density: femoral neck** Two uncontrolled, observational, retrospective studies provided evidence relating to the effect of gender-affirming hormones on femoral neck bone density in transfemales (Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results. **These studies provide very low certainty evidence that femoral neck bone density (measured by BMAD) was unchanged in sex assigned at birth males (transfemales) during treatment with gender-affirming hormones (follow-up between 2 and 5 years). Z-scores at the end of follow-up suggest and the average femoral neck bone density was lower than in the equivalent cisgender population. The results for femoral neck bone density (measured by BMD) were inconsistent.** **Change in clinical parameters: glucose, insulin and HbA1c** One uncontrolled, retrospective chart review (Klaver et al. 2020) provided evidence on glucose, insulin and HbA1c in transfemales. See the safety results table above for a full description of the results. **This study provided very low certainty evidence that gender-affirming hormones do not affect HbA1c, glucose levels, insulin levels and insulin resistance in sex assigned at birth males (transfemales) from the start of treatment to age 22 years.** **Change in clinical parameters: lipids** One retrospective chart review (Klaver et al. 2020) provided evidence on the change in lipids (total cholesterol, HDL cholesterol, LDL cholesterol and triglycerides) in transfemales. See the safety results table above for a full description of the results. **This study provides very low certainty evidence that gender-affirming hormones do not affect lipid profiles in sex assigned at birth males (transfemales) from the start of treatment to age 22 years.** **Change in clinical parameters: blood pressure** One retrospective chart review (Klaver et al. 2020) provided evidence on the change in blood pressure in transfemales. See the safety results table above for a full description of the results. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase blood pressure in sex assigned at birth males (transfemales),**

	although the absolute increase was small from the start of treatment to age 22 years. **Change in clinical parameters: body mass index (BMI)** One retrospective chart review (Klaver et al. 2020) provided evidence on the change in BMI in transfemales. See the safety results table above for a full description of the results. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase BMI in sex assigned at birth males (transfemales), although most participants were within the healthy weight range from the start of treatment to age 22 years.** **Treatment discontinuation** One uncontrolled, retrospective chart review provided evidence relating to permanent or temporary discontinuation of gender-affirming hormones in transfemales (Khatchadourian et al. 2014). **This study provides very low certainty evidence that the rates of discontinuation during treatment with gender-affirming hormones in sex assigned at birth males (transfemales) are low. Duration of treatment with gender-affirming hormones was not reported.** **Adverse effects** One uncontrolled, retrospective chart review provided evidence relating to adverse effects from gender-affirming hormones in transfemales (Khatchadourian et al. 2014). **This study provides very low certainty evidence about the potential adverse effects of gender-affirming hormones in sex assigned at birth males (transfemales). No conclusions could be drawn. Duration of treatment with gender-affirming hormones was not reported.**
Sex assigned at birth females (transmales) **Certainty of evidence: Very low**	Some studies reported data separately for sex assigned at birth females (transmales). This included some direct comparisons with sex assigned at birth males (transfemales). **Impact on mental health: depression and anxiety** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported the change in depression (measured using QIDS clinician-reported and self-reported), anxiety and anxiety-related symptoms (measured using SCARED) in transmales. See the clinical effectiveness results above for full details. In Kuper et al. 2020 (n=65 to 78, varies by outcome), changes were seen in depression, anxiety and anxiety-related symptoms from baseline to follow-up but the authors did not report any statistical analysis, so it is unclear if any changes are statistically significant **(VERY LOW)**.

	This study provides very low certainty evidence on the effects of gender-affirming hormones on depression, anxiety and anxiety-related symptoms over 10.9 months in transmales. No conclusions could be drawn. **Impact on mental health: suicidality** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in Ask Suicide-Screening Questions (ASQ) in transmales compared with transfemales. See the sex assigned at birth males (transfemales) row above for full details of the results. One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in suicidal ideation in transmales measured using additional questions from the PHQ 9_Modified for Teens. See the clinical effectiveness results above for full details. At baseline, 9.1% (3/33) of transmales had suicidal ideation, compared with 6.1% (2/33) at about 12-months follow-up (no statistical analysis reported) **(VERY LOW)**. **These studies provide very low certainty evidence that any change in suicidal ideation is not different between sex assigned at birth females (transmales) and sex assigned at birth males (transfemales). Mean duration of treatment about 12 months.** **Impact on quality of life** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in the GWBS of the Paediatric Quality of Life Inventory in transmales compared with transfemales. See the sex assigned at birth males (transfemales) row above for full details of the results. **This study provides very low certainty evidence that any change in general wellbeing is not different between sex assigned at birth females (transmales) and sex assigned at birth males (transfemales). Mean duration of treatment about 12 months.** **Impact on body image** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported change in Body Image Scale (BIS) in transmales. See the clinical effectiveness results above for full details. In Kuper et al. 2020 (n=66), the mean (±SD) BIS score was 71.1 points (±13.4) at baseline and 52.9 points (±16.8) at follow-up (no statistical analysis reported) **(VERY LOW)**. **This study provides very low certainty evidence on the effects of gender-affirming hormones on body image over 10.9 months in transmales. No conclusions could be drawn.** **Change in bone density: lumbar spine** Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of gender-affirming hormones on lumber spine bone density in transmales (Klink et al. 2015, Stoffers et al. 2019 and Vlot et al. 2017). See the safety results table above for a full details of the results.

	These studies provide very low certainty evidence that lumbar spine bone density (measured by BMAD) increases during 2 to 5 years treatment with gender-affirming hormones in sex assigned at birth females (transmales). Z-scores at the end of follow-up suggest the average lumbar spine bone density was generally lower than in the equivalent cisgender population. The results for lumbar spine bone density (measured by BMD) were inconsistent. **Change in bone density: femoral neck** Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of gender-affirming hormones on femoral neck bone density in transmales (Klink et al. 2015, Stoffers et al. 2019 and Vlot et al. 2017). See the safety results table above for a full details of the results. **These studies provide very low certainty evidence that femoral neck bone density (measured by BMAD) statistically significantly increased in sex assigned at birth females (transmales) during 2 to 5 years treatment with gender-affirming hormones. Z-scores at the end of follow-up suggest the average femoral neck bone density was generally lower than in the equivalent cisgender population. The results for femoral neck bone density (measured by BMD) were inconsistent.** **Change in clinical parameters: glucose, insulin and HbA1c** Two uncontrolled, retrospective chart reviews (Klaver et al. 2020; Stoffers et al. 2019) provided evidence on glucose, insulin and HbA1c in transmales. See the safety results table above for full details of the results. **This study provided very low certainty evidence that gender-affirming hormones do not affect HbA1c, glucose levels, insulin levels and insulin resistance in sex assigned at birth females (transmales). Reported from start of treatment to age 22 years.** **Change in clinical parameters: lipids** One retrospective chart review (Klaver et al. 2020) provided evidence on the change in lipids (total cholesterol, HDL cholesterol, LDL cholesterol and triglycerides) in transmales. See the safety results table above for full details of the results. **This study provides very low certainty evidence that treatment with gender-affirming hormones is associated with a small but statistically significant worsening of cholesterol levels in sex assigned at birth females (transmales), but mean cholesterol and triglyceride levels were within the UK reference range at end of treatment, from start of treatment to age 22 years.** **Change in clinical parameters: blood pressure** One retrospective chart review (Klaver et al. 2020) provided evidence on the change in blood pressure in transmales. See the safety results table above for full details of the results.

	This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase blood pressure in sex assigned at birth females (transmales), although the absolute increase was small, from start of treatment to age 22 years. **Change in clinical parameters: body mass index (BMI)** One retrospective chart review (Klaver et al. 2020) provided evidence on the change in body mass index (BMI) in transmales. See the safety results table above for full details of the results. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase BMI in sex assigned at birth females (transmales), although most participants were within the healthy weight range, from start of treatment to age 22 years.** **Change in clinical parameters: liver function** One retrospective chart review (Stoffers et al. 2019) provided non-comparative evidence on the change in liver enzymes in transmales between starting gender-affirming hormones and up to 24-months follow-up. See the safety results table above for full details of the results. **This study provides very low certainty evidence that gender-affirming hormones for about 12 months do not affect liver function in sex assigned at birth females (transmales).** **Change in clinical parameters: kidney function** One retrospective chart review (Stoffers et al. 2019) provided non-comparative evidence on the change in serum creatinine and serum urea levels in transmales between starting gender-affirming hormones and up to 24-months follow-up. See the safety results table above for full details of the results. **This study provides very low certainty evidence on the effects of gender-affirming hormones on kidney function in sex assigned at birth females (transmales). A statistically significant increase in creatinine levels was seen at about 12 months follow-up, but these were within the UK reference range. Urea levels were unchanged.** **Treatment discontinuation** One uncontrolled, retrospective chart review provided evidence relating to permanent or temporary discontinuation of gender-affirming hormones in transmales (Khatchadourian et al. 2014). See the safety results table above for full details of the results. **This study provides very low certainty evidence that the rates of treatment discontinuation with gender-affirming hormones in sex assigned at birth females (transmales) is low. Duration of gender-affirming hormones not reported.** **Adverse effects**

	One uncontrolled, retrospective chart review provided evidence for adverse effects from gender-affirming hormones in transmales (Khatchadourian et al. 2014). See the safety results table above for full details of the results. **This study provides very low certainty evidence about the potential adverse effects of gender-affirming hormones in sex assigned at birth females (transmales). No conclusions could be drawn. Duration of gender-affirming hormones not reported.**
Duration of gender dysphoria	No evidence was identified.
Age at onset of gender dysphoria	No evidence was identified.
Age at onset of puberty	No evidence was identified.
Tanner stage at which GnRH analogue or gender-affirming hormones started	One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported the impact of Tanner stage on outcomes, although it is not clear whether this is referring to Tanner stage at initial assessment, at the start of GnRH analogues or at another timepoint.
Diagnosis of autistic spectrum disorder	No evidence was identified.
Diagnosis of a mental health condition	One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported outcomes that were adjusted for engagement in counselling and medicines for mental health problems. Information about diagnoses and treatment were not provided. Rates of mental health issues appear to be high in the cohort. **Impact on mental health** Achille et al. 2020 reported the change in depression scores, controlled for engagement in counselling and medicines for mental health problems (measured using the Center for Epidemiologic Studies Depression [CESD-R] scale and Patient Health Questionnaire Modified for Teens [PHQ 9_Modified for Teens] score. • There was no statistically significant change in CESD-R from baseline to about 12-months follow-up. • There was no statistically significant change in PHQ 9_Modified for Teens score from baseline to about 12-months follow-up **(VERY LOW)**. **Impact on quality of life** Achille et al. 2020 reported the change in quality of life scores, controlled for engagement in counselling and medicines for mental health problems (measured using the Quality of Life Enjoyment and Satisfaction Questionnaire [QLES-Q-SF] score: • There was no statistically significant change in QLES-Q-SF score from baseline to about 12-months follow-up **(VERY LOW)**. **This study provides very low certainty evidence about outcomes that were adjusted for engagement in counselling and medicines for mental health problems. No conclusions could be drawn.**

Abbreviations: ASQ: Ask Suicide-Screening Questions; CESD-R: Center for Epidemiologic Studies Depression; GnRH: Gonadotrophin releasing hormone; GWBS: General Well-Being

Scale; HDL: high-density lipoproteins; LDL: low-density lipoproteins; p: p-value; PHQ 9_Modified for Teens: Patient Health Questionnaire Modified for Teens; QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire.

From the evidence selected,

(a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?

(b) what were the ages at which participants commenced treatment with gender-affirming hormones?

(c) what was the duration of treatment with GnRH analogues?

<table>
<tr><th>Outcome</th><th>Evidence statement</th></tr>
<tr><td>Diagnostic criteria</td><td>The DSM-IV-TR criteria was used in 3 studies (Klaver et al. 2020, Klink et al. 2015 and Vlot et al. 2017).

The DSM-V criteria was used in 2 studies (Kuper et al. 2020 and Stoffers et al. 2019). The DSM-V has one overarching definition of gender dysphoria with separate specific criteria for children and for adolescents and adults. The general definition describes a conflict associated with significant distress and/or problems functioning associated with this conflict between the way they feel and think of themselves which must have lasted at least 6 months.

The ICD-10 diagnosis of 'transsexualism' was used in 1 study (Kaltiala et al. 2020). The authors state that this is the corresponding diagnosis to 'gender dysphoria' in the DSM-V, and that diagnostic assessments in the study location (Finland) take place according to ICD-10.

It was not reported how gender dysphoria was defined in the remaining 4 studies (VERY LOW).

From the evidence selected, the most commonly reported diagnostic criteria for gender dysphoria (5/10 studies) was the DSM criteria in use at the time the study was conducted.</td></tr>
<tr><td>Age when gender-affirming hormones started</td><td>8/10 studies reported the age at which participants started treatment with gender-affirming hormones, either as the mean age (with SD) or median age (with the range):

<table>
<tr><th>Study</th><th>Mean age (± SD)</th></tr>
<tr><td>Allen et al. 2019</td><td>16.7 years (not reported)</td></tr>
<tr><td>Khatchadourian et al. 2014</td><td>17.4 years (1.9)</td></tr>
<tr><td>Klaver et al. 2020</td><td>16.4 years (1.1) in transfemales
16.9 years (0.9) in transmales</td></tr>
<tr><td>Kuper et al. 2020</td><td>16.2 (1.2)</td></tr>
<tr><td>Klink et al. 2015</td><td>16.6 years (1.4) in transfemales
16.4 years (2.3) in transmales</td></tr>
</table>

<table>
<tr><th>Study</th><th>Median age (range)</th></tr>
<tr><td>Stoffers et al. 2019</td><td>17.2 years (15 to 19.5)</td></tr>
<tr><td>Vlot et al. 2017</td><td>16.3 years (15.9 to 19.5) in transfemales
16.0 years (14.0 to 18.9) in transmales</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td>Age at the start of treatment was not reported in 3 studies:
In Achille et al. 2020 the mean age at initial assessment (baseline) was 16.2 years (SD ±2.2)
In Kaltiala et al. 2020 the mean age at diagnosis was 18.1 years (range 15.2 to 19.9)
In Lopez de Lara et al. 2020 the mean age of participants was 16 years (range 14 to 18), although it is not clear if this is at the initial assessment or at the start of gender-affirming hormones.
The evidence included showed that most children and adolescents started treatment with gender-affirming hormones at about 16 to 17 years, with a range of about 14 to 19 years.</td></tr>
<tr><td>Duration of treatment with GnRH analogues</td><td>The duration of treatment with GnRH analogues was reported in 3/10 studies:<table><tr><th>Study</th><th>Median duration</th></tr><tr><td>Klaver et al. 2020</td><td>2.1 years (IQR 1.0 to 2.7) in transfemales
1.0 years (IQR 0.5 to 2.9) in transmales</td></tr><tr><td>Klink et al. 2015</td><td>1.3 years (range 0.5 to 3.8) in transfemales
1.5 years (range 0.25 to 5.2) in transmales
(GnRH analogue monotherapy)</td></tr><tr><td>Stoffers et al. 2019</td><td>8 months (range 3 to 39)</td></tr></table>The evidence included showed wide variation in the duration of treatment with gender-affirming hormones, but most studies did not report this information. Treatment duration ranged from a few months up to about 5 years.</td></tr>
</table>

Abbreviations: DSM, Diagnostic and Statistical Manual of Mental Disorders criteria; GnRH, Gonadotrophin-releasing hormone; ICD, International Statistical Classification of Diseases and Related Health Problems; IQR, interquartile range; SD, standard deviation.

6. Discussion

A key limitation to identifying the effectiveness and safety of gender-affirming hormones for children and adolescents with gender dysphoria is the lack of reliable comparative studies. All the studies included in this evidence review are uncontrolled observational studies, which are subject to bias and confounding and were of very low certainty using modified GRADE. The size of the population with gender dysphoria means conducting a prospective trial may be unrealistic, at least on a single centre basis. There may also be ethical issues with a 'no treatment arm' in comparative trials of gender-affirming hormones, where there may be poor mental health outcomes if treatment is withheld. However, the use of an active comparator such as close psychological support may reduce ethical concerns in future trials. A fundamental limitation of all the uncontrolled studies included in this review is that any changes in scores from baseline to follow-up could be attributed to a regression-to-the-mean.

The included studies have relatively short follow-up, with an average duration of treatment with gender-affirming hormones between around 1 year and 5.8 years. Further studies with a

longer follow-up are needed to determine the long-term effect of gender-affirming hormones for children and adolescents with gender dysphoria.

Most studies included in this review did not report comorbidities (physical or mental health) and no study reported concomitant treatments in detail. Because of this it is not clear whether any changes observed were due to gender-affirming hormones or other treatments the participants may have received. For example, we do not know if any improvement in depression symptom score over time was the result of gender-affirming hormones or the mental health support the person may be receiving (including medicines or counselling). This may be of particular importance for the mental health outcomes discussed in this review, since depression, anxiety and other related symptoms are common in children and adolescents with gender dysphoria. In Achille et al. 2020, at baseline around one-third of participants were taking medicines for mental health problems and around two-thirds reported being depressed in the past year. In Kaltiala et al. 2020, half the participants needed mental health treatment during and before gender identity assessment, with the most common reasons for treatment being depression, anxiety and suicidality. Only 1 study reported outcomes adjusted for engagement in counselling and medicines for mental health problems (Achille et al. 2020). This study found that gender-affirming hormones had no significant impact on depression and quality of life when adjusted for mental health care, despite significant approvements reported for the unadjusted results. However, it is not possible to draw conclusions on the impact of concurrent mental health treatment on the effect of gender-affirming hormones based on this study alone. Details of the mental health care provided are not reported in the study and results are presented for transfemales and transmales separately, resulting in small patient numbers and possible underpowering.

In most of the included studies, details of the gender-affirming hormone treatment regimens are poorly reported, with limited information provided about the medicines, doses and routes of administration used. It is not clear whether the interventions used in the studies are reflective of current UK practice for children and adolescents with gender dysphoria. There is also the suggestion that the hormone dose used in 1 study may have been too low; the authors of Klink et al. 2015 suggest that the relatively low initial dose of oestrogen for transfemales may be the reason for the observed lack of effect on lumber spine bone density. Duration of treatment with a GnRH analogue is also poorly reported and is only stated in 3/10 studies.

There is a degree of indirectness in some studies, with some participants included that fall outside of the population of this evidence review. For example, in Kuper et al. 2020 17% of participants received puberty suppression alone, and in Achille et al. 2020, 30% of participants received no treatment or puberty suppression alone. Some results and statistical analyses are only reported for the whole cohort in these studies and not the subgroup of participants who received gender-affirming hormones.

Participant numbers are poorly reported in some of the included studies. In Achille et al. 2020, 47% (45/95) of the people who entered the study did not have follow-up data and were excluded from the analyses, with no explanation or description of those people lost to follow-up. In Kuper et al. 2020, the number of participants varied by outcome, with less than two-thirds of participants providing data for some outcomes. The authors provide no explanation for this incomplete reporting.

It is not clear whether some outcome measures, specifically those related to psychosocial functioning, are relevant to the UK population. In Kaltiala et al. 2020, an observational study conducted in Finland, the proportion of participants living with parents or guardians is reported as marker of appropriate functioning. The authors state that in Finnish culture young people tend to leave the parental home early, with only around one-quarter of 20 to 24 year olds still living at home. This is lower than in the UK, where around half of 20 to 24 year olds live with their parents or guardians (ONS: Why are more young people living with their parents?).

It is difficult to draw firm conclusions for many of the effectiveness and safety outcomes reported in the included studies because many different scoring tools and methods were used to assess the same outcome, often with conflicting results. For example, bone density is reported as bone mineral density (BMD) and bone mineral apparent density (BMAD) in the same study, the latter being a size-adjusted measure often useful for people whose bones are still growing. For some populations (transfemale versus transmale) and bone regions (lumber spine versus femoral neck), statistically significant differences in BMD are reported but not for BMAD, and vice versa.

In addition to this, most outcomes reported across the included studies do not have an accepted minimal clinically important difference (MCID), making it difficult the determine whether any observed statistically significant changes are clinically meaningful. However, the authors of some studies report thresholds to interpret the results of the scoring tools, so some conclusions can be made. For example, the mean Utrecht Gender Dysphoria Scale (UGDS) score (a measure of gender dysphoria symptoms) reduced to about 15 points after treatment with gender-affirming hormones (Lopez de Lara et al. 2020). The authors state that scores of 40 points or above signify gender dysphoria, suggesting that after about 12 months of treatment with gender-affirming hormones, the majority of participants did not have symptoms of gender dysphoria.

The impact of gender-affirming hormones on bone density was reported in 3 studies (Klink et al. 2015, Stoffers et al. 2019 and Vlot et al. 2017). Although these studies did not include a control group, comparisons to a reference population are reported using z-scores. Comparisons were made to a cisgender population, meaning for example that bone density in transfemales was compared with bone density in cisgender males. The authors of Klink et al. 2015 note that this may not be the ideal comparison, because androgens and oestrogens affect bone differently, and that bone properties in a trans population differ from their age- and sex assigned at birth-matched controls. Beyond this, a major limitation when trying to determine the impact of gender-affirming hormones on the short- and long-term bone health of children and adolescents is the lack of data on fracture rates and other patient-orientated outcomes, including rates of osteoporosis. Studies of GnRH analogues in children and adolescents with gender dysphoria suggest that GnRH analogue treatment may reduce the expected increase in bone density (which is expected during puberty). Although improvements in bone density were reported following treatment with gender-affirming hormones, Z-scores suggest that bone density remained lower in transfemales and transmales compared with an equivalent cisgender population.

One study reported on cardiovascular risk factors at age 22 years in people who started gender-affirming hormones for gender dysphoria as adolescents. While glucose levels, insulin levels and insulin resistance were broadly unchanged at 22 years, statistically

significant increases in blood pressure and body mass index were seen. A small but statistically significant worsening of the lipid profile in transmales who received testosterone was also seen at age 22 years. However, further studies with a considerably longer follow-up and a focus on patient-oriented outcomes, including cardiovascular events and mortality are needed to determine the long-term impact on cardiovascular health of starting gender-affirming hormones during childhood and adolescence.

Only 1 study reported adverse events and discontinuation rates with gender-affirming hormones in children and adolescents. Conclusions on these outcomes cannot be made based on this study alone.

This review did not identify sub-groups of people who may benefit more from gender-affirming hormones. Limited evidence from 2 studies suggests there was no difference in response to treatment between transfemales and transmales for mental health and quality of life (Achille et al. 2020 and Allen et al. 2019).

7. Conclusion

This evidence review found limited evidence for the effectiveness and safety of gender-affirming hormones in children and adolescents with gender dysphoria, with all studies being uncontrolled, observational studies, and all outcomes of very low certainty. Any potential benefits of treatment must be weighed against the largely unknown long-term safety profile of these treatments.

The results from 5 uncontrolled, observational studies (Achille et al. 2020, Allen et al. 2019, Kaltiala et al. 2020, Kuper et al. 2020, Lopez de Lara et al. 2020) suggest that, in children and adolescents with gender dysphoria, gender-affirming hormones are likely to improve symptoms of gender dysphoria, and may also improve depression, anxiety, quality of life, suicidality, and psychosocial functioning. The impact of treatment on body image is unclear. All results were of very low certainty. The clinical relevance of any improvements to the person is difficult to determine because most outcomes do not have a recognised minimal clinically important difference, and the authors do not present statistical analysis for some outcomes.

A further 5 uncontrolled, observational studies (Khatchadourian et al. 2014, Klaver et al. 2020, Klink et al. 2015, Stoffers et al. 2019 and Vlot et al. 2017) reported on safety outcomes, all of which provided very low certainty evidence. Statistically significant increases in some measures of bone density were seen following treatment with gender-affirming hormones, although results varied by bone region (lumber spine versus femoral neck) and by population (transfemales versus transmales). However, z-scores suggest that bone density remained lower in transfemales and transmales compared with an equivalent cisgender population. Results from 1 study of gender-affirming hormones started during adolescence reported statistically significant increases in blood pressure and body mass index, and worsening of the lipid profile (in transmales) at age 22 years, although longer term studies that report on cardiovascular event rates are needed. Adverse events and discontinuation rates associated with gender-affirming hormones were only reported in 1 study, and no conclusions can be made on these outcomes.

This review did not identify sub-groups of people who may benefit more from gender-affirming hormones. Limited evidence from 2 studies suggests there was no difference in

response to treatment between transfemales and transmales for mental health and quality of life (Achille et al. 2020 and Allen et al. 2019).

No cost-effectiveness evidence was found to determine whether gender-affirming hormones are a cost-effective treatment for children and adolescents with gender dysphoria.

Appendix A PICO

The review questions for this evidence review are:

1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
2. For children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
3. For children and adolescents with gender dysphoria, what is the cost-effectiveness of gender-affirming hormones compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?
4. From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria?
5. From the evidence selected,
   (a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?
   (b) what were the ages at which participants commenced treatment with gender-affirming hormones?
   (c) what was the duration of GnRH analogues treatment?

PICO table

P –Population and Indication	Children and adolescents aged 18 years or less who have gender dysphoria, gender identity disorder or gender incongruence of childhood as defined by the study. The following subgroups of children and adolescents with gender dysphoria, gender identity disorder or gender incongruence of childhood need to be considered: • Sex assigned at birth males • Sex assigned at birth females

	• The duration of gender dysphoria: less than 6 months, 6-24 months, and more than 24 months) • The age at which treatment was initiated with GnRH analogues and with gender-affirming hormones. • The age of onset of gender dysphoria • The age of onset of puberty • Adolescents with gender dysphoria who have a pre-existing diagnosis of autistic spectrum disorder. • Adolescents with gender dysphoria who had a significant mental health symptom load at diagnosis including anxiety, depression (with or without a history of self-harm and suicidality), psychosis, personality disorder, Attention Deficit Hyperactivity Disorder and eating disorders.
I – Intervention	Gender-affirming hormone treatments: • A testosterone preparation for sex assigned at birth female patients which may include testosterone in the form of Sustanon injections*; testosterone enantate injections; Tostran gel*; Testogel; Testim gel; oral testosterone capsules in the form of testosterone undecanoate (Restandol); Andriol testocaps; Nebido • An oestradiol preparation** for sex assigned at birth male patients which may include: oral estradiol valerate*; oestrogen patches (7β-oestradiol patches e.g. Evorel or Estradem); Estradot patches; ethinyloestradiol *** *These are the used by Leeds Hospital, England. ** Be aware that the American spelling is oestrogen without the 'o'. ***Ethinyloestradiol is rarely used.
C – Comparator(s)	One or a combination of: • Psychological support • Social transitioning to the gender with which the individual identifies. No intervention
O – Outcomes	There are no known minimal clinically important differences and there are no preferred timepoints for the outcome measures selected. **All outcomes should be stratified by:** • The age at which treatment with gender-affirming hormones was initiated • The length of treatment with GnRH analogues where possible. **A: Clinical Effectiveness** *Critical to decision making* • **Impact on gender dysphoria** This outcome is critical because gender dysphoria in adolescents and children is associated with significant distress and problems functioning. Impact on gender dysphoria may be measured by the Utrecht Gender

	Dysphoria Scale. Other measures as reported in studies may be used as an alternative to the stated measure. • **Impact on mental health** Examples of mental health problems include self-harm, thoughts of suicide, suicide attempts, suicide, eating disorders, depression/low mood and anxiety. These outcomes are critical because self-harm and thoughts of suicide have the potential to result in significant physical harm and for completed suicides the death of the young person. Disordered eating habits may cause significant morbidity in young people. Depression and anxiety are also critical outcomes because they may impact on social, occupational, or other areas of functioning of children and adolescents. The Child and Adolescent Psychiatric Assessment (CAPA) may be used to measure depression and anxiety. The impact on self-harm and suicidality (ideation and behaviour) may be measured using the Suicide Ideation Questionnaire Junior. Other measures may be used as an alternative to the stated measure. • **Impact on Quality of Life** This outcome is critical because gender dysphoria in children and adolescents may be associated with a significant reduction in health-related quality of life. Quality of Life may be measured by the KINDL questionnaire, Kidscreen 52. Other measures as reported in studies may be used as an alternative to the stated measures *Important to decision making* • **Impact on body image** This outcome is important because some young people with gender dysphoria may desire to take steps to suppress features of their physical appearance associated with their sex assigned at birth or accentuate physical features of their experienced gender. The Body Image Scale could be used as a measure. Other measures as reported in studies may also be used as an alternative to the stated measure. • **Psychosocial Impact** Examples of psychosocial impact are: coping mechanisms which may impact on substance misuse; family relationships; peer relationships. This outcome is important because gender dysphoria in adolescents and children is associated with internalising and externalising behaviours and emotional and behavioural problems which may impact on social and occupational functioning. The child behavioural check list (CBCL) may be used to measure the impact on psychosocial functioning. Other measures as reported in studies may be used as an alternative to the stated measure. • **Engagement with health care services** This outcome is important because patient engagement with healthcare services will impact on their clinical outcomes. Engagement with health care services may be measured using the Youth Health Care measure-satisfaction, utilization, and needs (YHC-SUN) questionnaire. Loss to follow up and should also be ascertained as part of this outcome.

	Alternative measures to the YHC-SUN questionnaire may be used as reported in studies. • **Transitioning surgery - Impact on extent of and satisfaction with surgery** This outcome is important because some children and adolescents with gender dysphoria may in adulthood proceed to transitioning surgery. Stated measures of the extent of surgery and satisfaction with surgery in studies may be reported. • **De-transition** The proportion of patients who de-transition following the commencement of gender-affirming hormone treatment and the reasons why. This outcome is important to patients because there is uncertainty about the short and long term safety and adverse effects of gender-affirming hormones in children and adolescents with gender dysphoria. **B: Safety** • Short and long -term safety and adverse effects of taking gender-affirming hormones is important to assess whether treatment causes acute side effects that may lead to withdrawing the treatment or long term effects that may impact on decisions for transitioning or de-transitioning. Aspects to be reported on should include Impact of the drug use such as clinically relevant derangement in renal and liver function tests, lipids, glucose, insulin and glycosylated haemoglobin, cognitive development and functioning. The clinical and physical impact of temporary and permanent withdrawal the drug such as when patients decide to de-transition – e.g. delay in the attainment of peak bone mass, attenuation of peak bone mass, permanent physical effects. **C: Cost effectiveness** Cost effectiveness studies should be reported.
Inclusion criteria	
Study design	Systematic reviews, randomised controlled trials, controlled clinical trials, cohort studies. If no higher level quality evidence is found, case series can be considered.
Language	English only
Patients	Human studies only
Age	18 years or less
Date limits	2000-2020

Exclusion criteria	
Publication type	Conference abstracts, non-systematic reviews, narrative reviews, commentaries, letters, editorials, guidelines and pre-publication prints
Study design	Case reports, resource utilisation studies

Appendix B Search strategy

Medline, Embase, the Cochrane Library, HTA and APA PsycInfo were searched on 21 July 2020, limiting the search to papers published in English language in the last 20 years. Conference abstracts, non-systematic reviews, narrative reviews, commentaries, letters, editorials, guidelines, pre-publication prints, case reports and resource utilisation studies were excluded.

Database: Medline
Platform: Ovid
Version: Ovid MEDLINE(R) <1946 to July 17, 2020>
Search date: 21 Jul 2020
Number of results retrieved: 650
Search strategy:
Database: Ovid MEDLINE(R) <1946 to July 17, 2020>
Search Strategy:
---

1 Gender Dysphoria/ (485)
2 Gender Identity/ (18431)
3 "Sexual and Gender Disorders"/ (75)
4 Transsexualism/ (3758)
5 Transgender Persons/ (3134)
6 Health Services for Transgender Persons/ (136)
7 exp Sex Reassignment Procedures/ (835)
8 (gender* adj3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (7223)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (12665)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (102312)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (6969)
12 (male-to-female or m2f or female-to-male or f2m).tw. (114785)
13 or/1-12 (252562)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (1137237)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (852126)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (1912796)
17 Minors/ (2572)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (2360626)
19 exp pediatrics/ (58102)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (835833)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (2023650)
22 Puberty/ (13277)

23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (424041)
24 Schools/ (38087)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (7199)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (468784)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (89314)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (887443)
29 or/14-28 (5532185)
30 13 and 29 (79220)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (7)
32 30 or 31 (79220)
33 Hormones/ad, tu, th (4514)
34 exp Progesterone/ad, tu, th (10899)
35 exp Estrogens/ad, tu, th (28936)
36 exp Gonadal Steroid Hormones/ad, tu, th (34137)
37 (progesteron* or oestrogen* or estrogen*).tw. (196074)
38 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (544)
39 exp Estradiol/ad, tu, th (10823)
40 exp Testosterone/ad, tu, th (8318)
41 (testosteron* or sustanon* or tostran or testogel or testim or restandol or androl or testocaps* or nebido or testavan).tw. (74936)
42 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (90464)
43 or/33-42 (304239)
44 32 and 43 (3183)
45 limit 44 to yr="2000 -Current" (2019)
46 animals/ not humans/ (4685420)
47 45 not 46 (1194)
48 limit 47 to english language (1155)
49 (MEDLINE or pubmed).tw. (163678)
50 systematic review.tw. (121198)
51 systematic review.pt. (130231)
52 meta-analysis.pt. (117148)
53 intervention$.ti. (123904)
54 or/49-53 (380217)
55 randomized controlled trial.pt. (509468)
56 randomi?ed.mp. (796957)
57 placebo.mp. (194937)
58 or/55-57 (848627)
59 exp cohort studies/ or exp epidemiologic studies/ or exp clinical trial/ or exp evaluation studies as topic/ or exp statistics as topic/ (5562241)
60 ((control and (group* or study)) or (time and factors)).mp. (3274107)
61 (program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. (4624419)
62 or/59-61 (9030680)
63 Observational Studies as Topic/ (5177)
64 Observational Study/ (81866)
65 Epidemiologic Studies/ (8358)

66 exp Case-Control Studies/ (1090891)
67 exp Cohort Studies/ (2011414)
68 Cross-Sectional Studies/ (332273)
69 Controlled Before-After Studies/ (526)
70 Historically Controlled Study/ (185)
71 Interrupted Time Series Analysis/ (913)
72 Comparative Study.pt. (1866044)
73 case control$.tw. (112152)
74 case series.tw. (59119)
75 (cohort adj (study or studies)).tw. (170281)
76 cohort analy$.tw. (6758)
77 (follow up adj (study or studies)).tw. (45131)
78 (observational adj (study or studies)).tw. (86247)
79 longitudinal.tw. (204239)
80 prospective.tw. (495367)
81 retrospective.tw. (442876)
82 cross sectional.tw. (284856)
83 or/63-82 (4368140)
84 54 or 58 or 62 or 83 (9402123)
85 48 and 84 (683)
86 limit 85 to (letter or historical article or comment or editorial or news or case reports) (33)
87 85 not 86 (650)

Database: Medline in-process
Platform: Ovid
Version: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <1946 to July 17, 2020>
Search date: 21 July 2020
Number of results retrieved: 122
Search strategy:
Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <1946 to July 17, 2020>
Search Strategy:
---

1 Gender Dysphoria/ (0)
2 Gender Identity/ (0)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (0)
5 Transgender Persons/ (0)
6 Health Services for Transgender Persons/ (0)
7 exp Sex Reassignment Procedures/ (0)
8 (gender* adj3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (1473)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (2315)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (20821)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (963)
12 (male-to-female or m2f or female-to-male or f2m).tw. (15453)
13 or/1-12 (39735)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (0)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (80295)

16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (0)
17 Minors/ (0)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (320315)
19 exp pediatrics/ (0)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (119124)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (0)
22 Puberty/ (0)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (59969)
24 Schools/ (0)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (0)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (68979)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (10287)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (112220)
29 or/14-28 (523053)
30 13 and 29 (9143)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw (3)
32 30 or 31 (9144)
33 Hormones/ad, tu, th (0)
34 exp Progesterone/ad, tu, th (0)
35 exp Estrogens/ad, tu, th (0)
36 exp Gonadal Steroid Hormones/ad, tu, th (0)
37 (progesteron* or oestrogen* or estrogen*).tw. (13291)
38 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (241)
39 exp Estradiol/ad, tu, th (0)
40 exp Testosterone/ad, tu, th (0)
41 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (5458)
42 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (4772)
43 or/33-42 (19706)
44 32 and 43 (316)
45 limit 44 to yr="2000 -Current" (303)
46 animals/ not humans/ (1)
47 45 not 46 (303)
48 limit 47 to english language (303)
49 (MEDLINE or pubmed).tw. (36030)
50 systematic review.tw. (29830)
51 systematic review.pt. (1007)
52 meta-analysis.pt. (49)
53 intervention$.ti. (21354)
54 or/49-53 (68976)
55 randomized controlled trial.pt. (277)
56 randomi?ed.mp. (74978)
57 placebo.mp. (18290)
58 or/55-57 (81427)
59 exp cohort studies/ or exp epidemiologic studies/ or exp clinical trial/ or exp evaluation studies as topic/ or exp statistics as topic/ (455)

60 ((control and (group* or study)) or (time and factors)).mp. (214372)
61 (program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. (339764)
62 or/59-61 (507046)
63 Observational Studies as Topic/ (0)
64 Observational Study/ (91)
65 Epidemiologic Studies/ (0)
66 exp Case-Control Studies/ (1)
67 exp Cohort Studies/ (1)
68 Cross-Sectional Studies/ (0)
69 Controlled Before-After Studies/ (0)
70 Historically Controlled Study/ (0)
71 Interrupted Time Series Analysis/ (0)
72 Comparative Study.pt. (46)
73 case control$.tw. (14451)
74 case series.tw. (13070)
75 (cohort adj (study or studies)).tw. (29119)
76 cohort analy$.tw. (1039)
77 (follow up adj (study or studies)).tw. (3540)
78 (observational adj (study or studies)).tw. (17421)
79 longitudinal.tw. (34485)
80 prospective.tw. (63689)
81 retrospective.tw. (73761)
82 cross sectional.tw. (60195)
83 or/63-82 (250805)
84 54 or 58 or 62 or 83 (687622)
85 48 and 84 (126)
86 limit 85 to (letter or historical article or comment or editorial or news or case reports) (4)
87 85 not 86 (122)

Database: Medline epubs ahead of print
Platform: Ovid
Version: Ovid MEDLINE(R) Epub Ahead of Print <July 17, 2020>
Search date: 21 July 2020
Number of results retrieved: 32
Search strategy:
Database: Ovid MEDLINE(R) Epub Ahead of Print <July 17, 2020>
Search Strategy:
---
1 Gender Dysphoria/ (0)
2 Gender Identity/ (0)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (0)
5 Transgender Persons/ (0)
6 Health Services for Transgender Persons/ (0)
7 exp Sex Reassignment Procedures/ (0)
8 (gender* adj3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (430)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (637)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (1499)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (179)
12 (male-to-female or m2f or female-to-male or f2m).tw. (2460)

13 or/1-12 (4883)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (0)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (15416)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (0)
17 Minors/ (0)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (53285)
19 exp pediatrics/ (0)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (22649)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (0)
22 Puberty/ (0)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (13005)
24 Schools/ (0)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (0)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (12420)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (1407)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (20083)
29 or/14-28 (87968)
30 13 and 29 (1618)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (1)
32 30 or 31 (1618)
33 Hormones/ad, tu, th (0)
34 exp Progesterone/ad, tu, th (0)
35 exp Estrogens/ad, tu, th (0)
36 exp Gonadal Steroid Hormones/ad, tu, th (0)
37 (progesteron* or oestrogen* or estrogen*).tw. (1876)
38 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (63)
39 exp Estradiol/ad, tu, th (0)
40 exp Testosterone/ad, tu, th (0)
41 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (846)
42 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (665)
43 or/33-42 (2850)
44 32 and 43 (64)
45 limit 44 to yr="2000 -Current" (61)
46 animals/ not humans/ (0)
47 45 not 46 (61)
48 limit 47 to english language (61)
49 (MEDLINE or pubmed).tw. (7948)
50 systematic review.tw. (7508)
51 systematic review.pt. (28)
52 meta-analysis.pt. (37)
53 intervention$.ti. (4267)
54 or/49-53 (15048)
55 randomized controlled trial.pt. (1)

56 randomi?ed.mp. (14113)
57 placebo.mp. (3097)
58 or/55-57 (15128)
59 exp cohort studies/ or exp epidemiologic studies/ or exp clinical trial/ or exp evaluation studies as topic/ or exp statistics as topic/ (34)
60 ((control and (group* or study)) or (time and factors)).mp. (31615)
61 (program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. (65735)
62 or/59-61 (88222)
63 Observational Studies as Topic/ (0)
64 Observational Study/ (4)
65 Epidemiologic Studies/ (0)
66 exp Case-Control Studies/ (0)
67 exp Cohort Studies/ (0)
68 Cross-Sectional Studies/ (0)
69 Controlled Before-After Studies/ (0)
70 Historically Controlled Study/ (0)
71 Interrupted Time Series Analysis/ (0)
72 Comparative Study.pt. (0)
73 case control$.tw. (2577)
74 case series.tw. (2480)
75 (cohort adj (study or studies)).tw. (7959)
76 cohort analy$.tw. (287)
77 (follow up adj (study or studies)).tw. (632)
78 (observational adj (study or studies)).tw. (3763)
79 longitudinal.tw. (7079)
80 prospective.tw. (12148)
81 retrospective.tw. (16600)
82 cross sectional.tw. (9459)
83 or/63-82 (48534)
84 54 or 58 or 62 or 83 (119752)
85 48 and 84 (32)
86 limit 85 to (letter or historical article or comment or editorial or news or case reports) (0)
87 85 not 86 (32)

Database: Medline daily update
Platform: Ovid
Version: Ovid MEDLINE(R) Daily Update <July 21, 2020>
Search date: 22 July 2020
Number of results retrieved: 3
Search strategy

Database: Ovid MEDLINE(R) Daily Update <July 21, 2020>
Search Strategy:
---

1 Gender Dysphoria/ (4)
2 Gender Identity/ (38)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (2)
5 Transgender Persons/ (26)
6 Health Services for Transgender Persons/ (1)
7 exp Sex Reassignment Procedures/ (3)
8 (gender* adj3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (22)

9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (39)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (87)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (15)
12 (male-to-female or m2f or female-to-male or f2m).tw. (181)
13 or/1-12 (358)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (932)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (981)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (1756)
17 Minors/ (3)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (3672)
19 exp pediatrics/ (75)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (1658)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (2006)
22 Puberty/ (8)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (732)
24 Schools/ (56)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (5)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (622)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (98)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (1301)
29 or/14-28 (6705)
30 13 and 29 (130)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (0)
32 30 or 31 (130)
33 Hormones/ad, tu, th (3)
34 exp Progesterone/ad, tu, th (3)
35 exp Estrogens/ad, tu, th (8)
36 exp Gonadal Steroid Hormones/ad, tu, th (22)
37 (progesteron* or oestrogen* or estrogen*).tw. (161)
38 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (3)
39 exp Estradiol/ad, tu, th (8)
40 exp Testosterone/ad, tu, th (8)
41 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (79)
42 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (61)
43 or/33-42 (261)
44 32 and 43 (7)
45 limit 44 to yr="2000 -Current" (7)
46 animals/ not humans/ (3647)
47 45 not 46 (6)
48 limit 47 to english language (6)
49 (MEDLINE or pubmed).tw. (529)
50 systematic review.tw. (512)

51 systematic review.pt. (522)
52 meta-analysis.pt. (370)
53 intervention$.ti. (247)
54 or/49-53 (1065)
55 randomized controlled trial.pt. (595)
56 randomi?ed.mp. (1203)
57 placebo.mp. (219)
58 or/55-57 (1234)
59 exp cohort studies/ or exp epidemiologic studies/ or exp clinical trial/ or exp evaluation studies as topic/ or exp statistics as topic/ (7958)
60 ((control and (group* or study)) or (time and factors)).mp. (4307)
61 (program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. (5828)
62 or/59-61 (11814)
63 Observational Studies as Topic/ (27)
64 Observational Study/ (449)
65 Epidemiologic Studies/ (7)
66 exp Case-Control Studies/ (2173)
67 exp Cohort Studies/ (3287)
68 Cross-Sectional Studies/ (837)
69 Controlled Before-After Studies/ (1)
70 Historically Controlled Study/ (0)
71 Interrupted Time Series Analysis/ (6)
72 Comparative Study.pt. (768)
73 case control$.tw. (182)
74 case series.tw. (139)
75 (cohort adj (study or studies)).tw. (561)
76 cohort analy$.tw. (22)
77 (follow up adj (study or studies)).tw. (40)
78 (observational adj (study or studies)).tw. (253)
79 longitudinal.tw. (429)
80 prospective.tw. (778)
81 retrospective.tw. (1032)
82 cross sectional.tw. (739)
83 or/63-82 (5471)
84 54 or 58 or 62 or 83 (12581)
85 48 and 84 (3)
86 limit 85 to (letter or historical article or comment or editorial or news or case reports) (0)
87 85 not 86 (3)

Database: Embase
Platform: Ovid
Version: Embase <1974 to 2020 July 22>
Search date: 23 July 2020
Number of results retrieved: 1207
Search strategy:

Database: Embase <1974 to 2020 July 22>
Search Strategy:
---
1 exp Gender Dysphoria/ (5399)
2 Gender Identity/ (16820)
3 "Sexual and Gender Disorders"/ (24689)
4 Transsexualism/ (3869)
5 exp Transgender/ (6597)

6 Health Services for Transgender Persons/ (158848)
7 exp Sex Reassignment Procedures/ (1108)
8 (gender* adj3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (12470)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (22509)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (154446)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (10327)
12 (male-to-female or m2f or female-to-male or f2m).tw. (200166)
13 or/1-12 (581748)
14 exp juvenile/ or Child Behavior/ or Child Welfare/ or Child Health/ or infant welfare/ or "minor (person)"/ or elementary student/ or adolescent health/ or middle school student/ or high school student/ (3440943)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (1186161)
16 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (3586795)
17 exp pediatrics/ (106214)
18 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (1491597)
19 exp adolescence/ or exp adolescent behavior/ or adolescent health/ or high school student/ or middle school student/ (105108)
20 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (641660)
21 school/ or high school/ or kindergarten/ or middle school/ or primary school/ or nursery school/ or day care/ (103791)
22 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (687437)
23 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (138908)
24 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (1562903)
25 or/14-24 (7130881)
26 13 and 25 (181778)
27 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (17)
28 26 or 27 (181778)
29 hormone/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (5160)
30 exp progesterone derivative/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (23479)
31 exp estrogen/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (57641)
32 steroid hormone/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (372)
33 sex hormone/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (1984)
34 hormonal therapy/ (42222)
35 (progesteron* or oestrogen* or estrogen*).tw. (254142)
36 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (1224)
37 exp estradiol derivative/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (30740)

38 exp testosterone derivative/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (15868)
39 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (99596)
40 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (114290)
41 or/29-40 (438737)
42 28 and 41 (6053)
43 limit 42 to yr="2000 -Current" (4741)
44 nonhuman/ not human/ (4649157)
45 43 not 44 (3636)
46 limit 45 to english language (3513)
47 (MEDLINE or pubmed).tw. (261145)
48 exp systematic review/ or systematic review.tw. (302985)
49 meta-analysis/ (191173)
50 intervention$.ti. (200041)
51 or/47-50 (660206)
52 random:.tw. (1552336)
53 placebo:.mp. (455979)
54 double-blind:.tw. (210671)
55 or/52-54 (1807280)
56 cohort analysis/ (596360)
57 exp epidemiology/ (3434332)
58 exp clinical trial/ (1504711)
59 evaluation study/ (45870)
60 statistics/ (301181)
61 ((control and (group* or study)) or (time and factors)).mp. (3324555)
62 (program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. (6067112)
63 or/56-62 (11048972)
64 Clinical study/ (155444)
65 Case control study/ (157943)
66 Family study/ (26047)
67 Longitudinal study/ (141660)
68 Retrospective study/ (937696)
69 comparative study/ (859061)
70 Prospective study/ (613138)
71 Randomized controlled trials/ (182542)
72 70 not 71 (606604)
73 Cohort analysis/ (596360)
74 cohort analy$.tw. (13020)
75 (Cohort adj (study or studies)).tw. (302159)
76 (Case control$ adj (study or studies)).tw. (137432)
77 (follow up adj (study or studies)).tw. (63423)
78 (observational adj (study or studies)).tw. (168428)
79 (epidemiologic$ adj (study or studies)).tw. (106448)
80 (cross sectional adj (study or studies)).tw. (220073)
81 case series.tw. (104089)
82 prospective.tw. (861922)
83 retrospective.tw. (886445)
84 or/64-69,72-83 (4047788)
85 51 or 55 or 63 or 84 (12494560)
86 46 and 85 (2151)
87 86 not (letter or editorial).pt. (2137)

88 87 not (conference abstract or conference paper or conference proceeding or "conference review").pt. (1207)

Database: APA PsycInfo

Platform: Ovid
Version: APA PsycInfo <1806 to July Week 2 2020>
Search date: 22 July 2020
Number of results retrieved: 581
Search strategy:

Database: APA PsycInfo <1806 to July Week 2 2020>
Search Strategy:
---

1 Gender Dysphoria/ (936)
2 Gender Identity/ (8648)
3 Transsexualism/ (2825)
4 Transgender/ (5257)
5 exp Gender Reassignment/ (568)
6 (gender* adj3 (dysphori* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (15276)
7 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (13028)
8 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (7679)
9 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (5796)
10 (male-to-female or m2f or female-to-male or f2m).tw. (63688)
11 or/1-10 (99498)
12 exp Infant Development/ (21841)
13 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (150219)
14 Child Characteristics/ or exp Child Behavior/ or Child Psychology/ or exp Child Welfare/ or Child Psychiatry/ (23423)
15 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (984230)
16 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (78962)
17 Adolescent Psychiatry/ or Adolescent Behavior/ or Adolescent Development/ or Adolescent Psychology/ or Adolescent Characteristics/ or Adolescent Health/ (62142)
18 Puberty/ (2753)
19 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (347604)
20 Schools/ (29181)
21 Child Day Care/ or Nursery Schools/ (2836)
22 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (772814)
23 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (21475)
24 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (285697)
25 or/12-24 (1765408)
26 11 and 25 (49560)
27 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw (14)

28 26 or 27 (49561)
29 hormones/ (8408)
30 sex hormones/ (1777)
31 exp progestational hormones/ (2409)
32 estrogens/ (3889)
33 steroids/ (3797)
34 (progesteron* or oestrogen* or estrogen*).tw. (11188)
35 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (457)
36 estradiol/ (3120)
37 testosterone/ (5606)
38 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (9625)
39 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (6741)
40 or/29-39 (30344)
41 28 and 40 (1005)
42 limit 41 to yr="2000 -Current" (749)
43 limit 42 to english language (692)
44 limit 43 to ("0200 book" or "0240 authored book" or "0280 edited book" or "0300 encyclopedia" or "0400 dissertation abstract") (111)
45 43 not 44 (581)

Database: Cochrane Library – incorporating Cochrane Database of Systematic Reviews (CDSR); CENTRAL

Platform: Wiley
Version:
CDSR –Issue 7 of 12, July 2020
CENTRAL – Issue 7 of 12, July 2020
Search date: 22 July 2020
Number of results retrieved: CDSR 0 ; CENTRAL 67

ID Search Hits
#1 MeSH descriptor: [Gender Dysphoria] this term only 3
#2 MeSH descriptor: [Gender Identity] this term only 227
#3 MeSH descriptor: [Sexual and Gender Disorders] this term only 2
#4 MeSH descriptor: [Transsexualism] this term only 27
#5 MeSH descriptor: [Transgender Persons] this term only 36
#6 MeSH descriptor: [Health Services for Transgender Persons] this term only 0
#7 MeSH descriptor: [Sex Reassignment Procedures] explode all trees 4
#8 (gender* near/3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)):ti,ab,kw 702
#9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*):ti,ab,kw 959
#10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*):ti,ab,kw 3969
#11 ((sex or gender*) near/3 (reassign* or chang* or transform* or transition*)):ti,ab,kw 524
#12 (male-to-female or m2f or female-to-male or f2m):ti,ab,kw 516
#13 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 6413
#14 MeSH descriptor: [Infant] explode all trees 28440
#15 MeSH descriptor: [Infant Health] this term only 49
#16 MeSH descriptor: [Infant Welfare] this term only 82

#17 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*):ti,ab,kw,so 89530
#18 MeSH descriptor: [Child] explode all trees 44089
#19 MeSH descriptor: [Child Behavior] explode all trees 2061
#20 MeSH descriptor: [Child Health] this term only 98
#21 MeSH descriptor: [Child Welfare] this term only 325
#22 MeSH descriptor: [Minors] this term only 8
#23 (child* or minor or minors or boy* or girl* or kid or kids or young*):ti,ab,kw,so 265417
#24 MeSH descriptor: [Pediatrics] explode all trees 661
#25 (pediatric* or paediatric* or peadiatric*):ti,ab,kw,so 57725
#26 MeSH descriptor: [Adolescent] this term only 102154
#27 MeSH descriptor: [Adolescent Behavior] this term only 1358
#28 MeSH descriptor: [Adolescent Health] this term only 29
#29 MeSH descriptor: [Puberty] this term only 295
#30 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*):ti,ab,kw,so 140927
#31 MeSH descriptor: [Schools] this term only 1914
#32 MeSH descriptor: [Child Day Care Centers] this term only 231
#33 MeSH descriptor: [Nurseries, Infant] explode all trees 17
#34 MeSH descriptor: [Schools, Nursery] this term only 37
#35 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*):ti,ab,kw,so 97810
#36 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") near/2 (year or years or age or ages or aged)):ti,ab 6710
#37 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") near/2 (year or years or age or ages or aged)):ti,ab 196881
#38 #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 516067
#39 #13 and #38 2488
#40 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*):ti,ab,kw 0
#41 #39 or #40 2488
#42 MeSH descriptor: [Hormones] this term only 2241
#43 MeSH descriptor: [Progesterone] explode all trees 3135
#44 MeSH descriptor: [Estrogens] explode all trees 1841
#45 MeSH descriptor: [Gonadal Steroid Hormones] explode all trees 10747
#46 (progesteron* or oestrogen* or estrogen*):ti,ab,kw 18387
#47 ((cross-sex or crosssex or gender-affirm*) and (hormon* or sterold* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)):ti,ab,kw 24
#48 MeSH descriptor: [Estradiol] explode all trees 4434
#49 MeSH descriptor: [Testosterone] explode all trees 2945
#50 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan):ti,ab,kw 7386
#51 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle):ti,ab,kw 11410
#52 #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 31870
#53 #41 and #52 121
#54 "conference":pt or (clinicaltrials or trialsearch):so 492465
#55 #53 not #54 72

Database: HTA

Platform: Wiley
Version: up to 2018
Search date: 22nd July 2020
Number of results retrieved: 4
Search strategy:

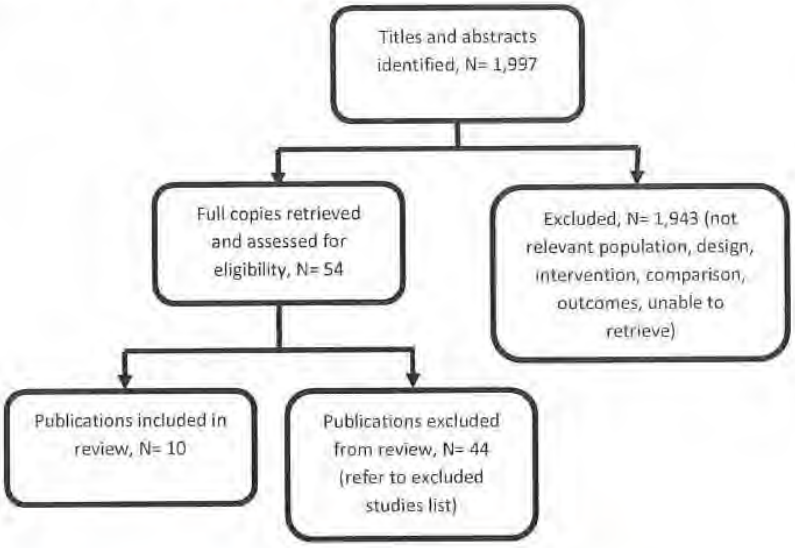
#1 MeSH DESCRIPTOR Gender Dysphoria 0
#2 MeSH DESCRIPTOR Gender Identity 12
#3 MeSH DESCRIPTOR Sexual and Gender Disorders 2
#4 MeSH DESCRIPTOR Transsexualism 12
#5 MeSH DESCRIPTOR Transgender Persons 3
#6 MeSH DESCRIPTOR Health Services for Transgender Persons 0
#7 MeSH DESCRIPTOR Sex Reassignment Procedures EXPLODE ALL TREES 1
#8 ((gender* near3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*))) 28
#9 ((transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*)) 76
#10 ((trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*)) 83
#11 (((sex or gender*) near3 (reassign* or chang* or transform* or transition*))) 24
#12 ((male-to-female or m2f or female-to-male or f2m)) 86
#13 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 261
#14 MeSH DESCRIPTOR Infant EXPLODE ALL TREES 2964
#15 MeSH DESCRIPTOR Infant Health 0
#16 MeSH DESCRIPTOR Infant Welfare 22
#17 ((prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*)) 5510
#18 MeSH DESCRIPTOR Child EXPLODE ALL TREES 4935
#19 MeSH DESCRIPTOR Child Behavior EXPLODE ALL TREES 64
#20 MeSH DESCRIPTOR Child Health 2
#21 MeSH DESCRIPTOR Child Welfare 80
#22 MeSH DESCRIPTOR Minors 2
#23 ((child* or minor or minors or boy* or girl* or kid or kids or young*)) 13575
#24 MeSH DESCRIPTOR Pediatrics EXPLODE ALL TREES 119
#25 ((pediatric* or paediatric* or peadiatric*)) 2842
#26 MeSH DESCRIPTOR Adolescent 4594
#27 MeSH DESCRIPTOR Adolescent Behavior 94
#28 MeSH DESCRIPTOR Adolescent Health 0
#29 MeSH DESCRIPTOR Puberty 3
#30 ((adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*)) 5621
#31 MeSH DESCRIPTOR Schools 168
#32 MeSH DESCRIPTOR Child Day Care Centers 12
#33 MeSH DESCRIPTOR Schools, Nursery 3
#34 ((pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*)) 4454
#35 ((("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") near2 (year or years or age or ages or aged))) 380
#36 ((("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") near2 (year or years or age or ages or aged)))7996

#37 #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 22640
#38 #13 AND #37 116
#39 (#13 AND #37) IN HTA 4

Appendix C Evidence selection

The literature searches identified 1,997 references. These were screened using their titles and abstracts and 54 references were obtained and assessed for relevance. Of these, 10 references are included in the evidence review. The remaining 44 references were excluded and are listed in appendix D.

Figure 1 – Study selection flow diagram



References submitted with Preliminary Policy Proposal

There is no preliminary policy proposal for this policy.

Appendix D Excluded studies table

Study reference	Reason for exclusion
Aranda G, Mora M, Hanzu FA et al. (2019) Effects of sex steroids on cardiovascular risk profile in transgender men under gender affirming hormone therapy. Endocrinologia, diabetes y nutricion 66(6): 385–392	Excluded on population – adult study, participants not 18 years or less (mean age 27.1 years).
Arnold, Justin D, Sarkodie, Eleanor P, Coleman, Megan E et al. (2016) Incidence of Venous Thromboembolism in Transgender Women	Excluded on population – adult study, participants not 18 years or less (mean age 33.2 years).

Study reference	Reason for exclusion
Receiving Oral Estradiol. The journal of sexual medicine 13(11): 1773–1777	
Asscheman, Henk, Giltay, Erik J, Megens, Jos A J et al. (2011) A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones. European journal of endocrinology 164(4): 635–42	Excluded on population – although some participants started gender-affirming hormones when young, the study does not report the proportion who started treatment when 18 years or less. Mean ages at start of treatment were 31.4 years (transfemales) and 26.1 years (transmales), suggesting the majority of participants were older than 18 years at the start of treatment. Outcomes not reported separately for people aged 18 years or less.
Author not, found (2014) Hormone therapy for the treatment of gender dysphoria. Lansdale, PA: HAYES, Inc	Full text paper not available.
Baba, T., Endo, T., Honnma, H. et al. (2007) Association between polycystic ovary syndrome and female-to-male transsexuality. Human Reproduction 22(4): 1011–1016	Excluded on population – although study included some younger people (age range 17 to 47), most participants were adults (mean age around 25 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.
Becerra-Fernandez A, Perez-Lopez G, Roman MM et al. (2014) Prevalence of hyperandrogenism and polycystic ovary syndrome in female to male transsexuals. Endocrinologia y Nutricion: Organo de la Sociedad Espanola de Endocrinologia y Nutricion 61(7): 351–8	Excluded on population – although study included some younger people (age range 18 to 45), most participants were adults (mean age around 25 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.
Becker I, Auer M, Barkmann C et al. (2018) A Cross-Sectional Multicenter Study of Multidimensional Body Image in Adolescents and Adults with Gender Dysphoria Before and After Transition-Related Medical Interventions. Archives of Sexual Behavior 47(8): 2335–2347	Excluded on population – study included people aged 14 to 21 years. Outcomes not reported separately for people aged 18 years or less. Better evidence available – only 11 participants received gender-affirming hormones. The majority of the study cohort were either pre-treatment, received puberty suppression alone, or received hormones and underwent surgery
Chew D, Anderson J, Williams K et al. (2018) Hormonal Treatment in Young People With Gender Dysphoria: A Systematic Review. Pediatrics 141(4): e20173742	Excluded on better available evidence - systematic review did not meta-analyse results from. Individual studies from this systematic review are either

Study reference	Reason for exclusion
	included, or excluded because they did not meet the PICO criteria.
Connolly MD, Zervos MJ, Barone CJ 2nd et al. (2016) The Mental Health of Transgender Youth: Advances in Understanding. The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine 59(5): 489–495	Excluded on intervention - review did not investigate gender-affirming hormones
de Vries ALC, McGuire JK, Steensma TD et al. (2014) Young adult psychological outcome after puberty suppression and gender reassignment. Pediatrics 134(4): 696–704	Exclude on intervention – all participants had surgery after gender-affirming hormones. Unable to determine whether changes were due to hormones or surgery. Complete data only available for 40 patients. Details of gender-affirming hormones are poorly reported. Outcomes reported in other study (with a population that more closely matches PICO)
Elamin MB, Garcia MZ, Murad MH et al. (2010) Effect of sex steroid use on cardiovascular risk in transsexual individuals: a systematic review and meta-analyses. Clinical Endocrinology 72(1): 1–10	Exclude on population – all included studies conducted in adult population. Unclear whether hormones were started when participants were aged 18 years or less. Outcomes not reported by age at treatment initiation.
Fernandez JD and Tannock LR (2016) Metabolic effects of hormone therapy in transgender patients. Endocrine Practice: Official Journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 22(4): 383–8	Excluded on population – adult study, participants not 18 years or less (mean ages 31 and 27 years).
Fighera TM, Ziegelmann PK, Da Silva TR et al. (2019) Bone mass effects of cross-sex hormone therapy in transgender people: Updated systematic review and meta-analysis. Journal of the Endocrine Society 3(5): 943–964	Excluded on population – all included studies conducted in adult population. Unclear whether hormones were started when participants were aged 18 years or less. Outcomes not reported by age at treatment initiation.
Getahun D, Nash R, Flanders WD et al. (2018) Cross-sex Hormones and Acute Cardiovascular Events in Transgender Persons: A Cohort Study. Annals of Internal Medicine 169(4): 205–213	Excluded on population – adult study, participants not 18 years or less
Gomez-Gil E, Zubiaurre-Elorza L, de Antonio IE et al. (2014) Determinants of quality of life in Spanish transsexuals attending a gender unit before genital sex reassignment surgery. Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation 23(2): 669–76	Excluded on population – although study included some younger people (age range 16 to 67), most participants were adults (mean age 31.2 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.
Gomez-Gil E, Zubiaurre-Elorza L, Esteva I et al. (2012) Hormone-treated transsexuals report less	Excluded on population – adult study, participants not 18 years or less (mean age 24.6 years).

Study reference	Reason for exclusion
social distress, anxiety and depression. Psychoneuroendocrinology 37(5): 662–70	
Gooren LJ, van Trotsenburg MAA, Giltay EJ et al. (2013) Breast cancer development in transsexual subjects receiving cross-sex hormone treatment. The Journal of Sexual Medicine 10(12): 3129–34	Excluded on population – study reports on cancer rates in people aged 18-80 years. The 3 cases of cancer all started gender-affirming hormone treatment >18 years.
Grimstad FW, Boskey E, Grey M (2020) New-Onset Abdominopelvic Pain After Initiation of Testosterone Therapy Among TransMasculine Persons: A Community-Based Exploratory Survey. LGBT health 7(5): Published Online:13 Jul 2020https://doi.org/10.1089/lgbt.2019.0258	Excluded on population – adult study, participants not 18 years or less.
Hannema SE, Schagen SEE, Cohen-Kettenis PT et al. (2017) Efficacy and Safety of Pubertal Induction Using 17beta-Estradiol in Transgirls. The Journal of Clinical Endocrinology and Metabolism 102(7): 2356–2363	Excluded on better evidence available – small study (n=28) with high drop-out rate (n=16 at final follow-up). Same outcomes reported in larger studies.
Jarin J, Pine-Twaddell E, Trotman G et al. (2017) Cross-Sex Hormones and Metabolic Parameters in Adolescents With Gender Dysphoria. Pediatrics 139(5)	Excluded on population and better evidence available. Although the study included some younger people (age range 13 to 25; mean age 16 and 18), the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less. Outcomes were limited to physiological results (including haemoglobin, lipids and BMI). Follow-up only 6 months, other included studies report same outcomes with longer follow-up (12 to 31 months).
Keo-Meier CL, Herman LI, Reisner SL et al. (2015) Testosterone treatment and MMPI-2 improvement in transgender men: a prospective controlled study. Journal of consulting and clinical psychology 83(1): 143–56	Excluded on population – although study included some younger people (age range 18 to 54), most participants were adults (mean age 26.6 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.
Klaver M, de Mutsert R, Wiepjes CM et al. (2018) Early Hormonal Treatment Affects Body Composition and Body Shape in Young Transgender Adolescents. The Journal of Sexual Medicine 15(2): 251–260	Excluded on outcomes – reported outcomes not included in PICO document. The risk of obesity with gender-affirmed hormones was reported in an included study.
McFarlane T, Zajac JD, Cheung AS (2018) Gender-affirming hormone therapy and the risk of sex hormone-dependent tumours in transgender individuals-A systematic review. Clinical Endocrinology 89(6): 700-711	Exclude on population – all included studies conducted in adult population.

Study reference	Reason for exclusion
Meriggiola MC, Armillotta F, Costantino A et al. (2008) Effects of testosterone undecanoate administered alone or in combination with letrozole or dutasteride in female to male transsexuals. The Journal of Sexual Medicine 5(10): 2442–53	Excluded on population – adult study, participants not 18 years or less.
Nota NM, Wiepjes CM, de Blok, CJM et al. (2018) The occurrence of benign brain tumours in transgender individuals during cross-sex hormone treatment. Brain: A Journal of Neurology 141(7): 2047–2054	Excluded on population – adult study, participants not 18 years or less.
Oda H and Kinoshita T (2017) Efficacy of hormonal and mental treatments with MMPI in FtM individuals: Cross-sectional and longitudinal studies. BMC Psychiatry 17(1): 256	Excluded on population – although study included some younger people (age range 15 to 43), most participants were adults (mean age around 25.6 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.
Olson-Kennedy J, Okonta V, Clark LF et al. (2018) Physiologic Response to Gender-Affirming Hormones Among Transgender Youth. The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine 62(4): 397–401	Excluded on population – although study included some younger people (age range 12 to 23; mean age 18 years). Outcomes not reported separately for people aged 18 years or less. Outcomes limited to physiological results (including haemoglobin, lipids, liver enzymes and BMI). Same outcomes reported in included studies that had a less indirect population and a longer follow-up.
Ott J, Kaufmann U, Bentz K et al. (2010) Incidence of thrombophilia and venous thrombosis in transsexuals under cross-sex hormone therapy. Fertility and sterility 93(4): 1267–72	Excluded on population – adult study, participants not 18 years or less.
Pakpoor J, Wotton CJ, Schmierer K et al. (2016) Gender identity disorders and multiple sclerosis risk: A national record-linkage study. Multiple sclerosis. Multiple Sclerosis Journal. 22(13): 1759–1762	Excluded on population – although study included some younger people, outcomes not reported separately for people aged 18 years or less. Also exclude for intervention – unclear if people received gender-affirming hormones.
Pyra M, Casimiro I, Rusie L et al. (2020) An Observational Study of Hypertension and Thromboembolism among Transgender Patients Using Gender-Affirming Hormone Therapy. Transgender Health 5(1): 1–9	Excluded on population – adult study (age range 20-70). Age at which gender-affirming hormones started not reported.
Quiros C, Patrascioiu I, Mora M et al. (2015) Effect of cross-sex hormone treatment on cardiovascular risk factors in transsexual individuals. Experience in a specialized unit in Catalonia. Endocrinologia y nutricion : organo de la Sociedad Espanola de Endocrinologia y Nutricion 62(5): 210–6	Excluded on population – adult study, participants not 18 years or less.

Study reference	Reason for exclusion
Rowniak S, Bolt L, Sharifi C (2019) Effect of cross-sex hormones on the quality of life, depression and anxiety of transgender individuals: A quantitative systematic review. JBI Database of Systematic Reviews and Implementation Reports 17(9): 1826–1854	Exclude on population – all included studies conducted in adult population.
Sequeira GM, Kidd K, El Nokali NE et al. (2019) Early Effects of Testosterone Initiation on Body Mass Index in Transmasculine Adolescents. Journal of Adolescent Health 65(6): 818–820	Exclude on outcome - study only reports BMI z-score over 12 month testosterone treatment. BMI not listed as an outcome of interest in the PICO document. Other included studies have investigated the impact of gender-affirming hormone treatment on CV risk profile, including longer term obesity rates, with a longer follow-up and more participants.
Shim JY, Laufer MR, Grimstad FW (2020) Dysmenorrhea and Endometriosis in Transgender Adolescents. Journal of Pediatric and Adolescent Gynecology. Available online 11 June 2020. https://doi.org/10.1016/j.jpag.2020.06.001	Exclude on population – only 2 participants taking testosterone before diagnosis of dysmenorrhea.
Slabbekoorn D, Van Goozen SHM, Gooren, LJG et al. (2001) Effects of cross-sex hormone treatment on emotionality in transsexuals. International Journal of Transgenderism 5(3): http://www.symposion.com/ijt/ijtvo05no03_02.htm	Excluded on population – adult study (age range 21 to 28 years)
Smith YLS., Van Goozen SHM, Kuiper AJ et al. (2005) Sex reassignment. Outcomes and predictors of treatment for adolescent and adult transsexuals. Psychological Medicine 35(1): 89–99	Excluded on population – results on adults only used to assess hormone treatment.
Sutherland N, Espinel W, Grotzke M et al. (2020) Unanswered Questions: Hereditary breast and gynecological cancer risk assessment in transgender adolescents and young adults. Journal of Genetic Counseling 29(4): 625–633	Excluded on study type – narrative review of 3 case reports.
van Velzen DM, Paldino A, Klaver M et al. (2019) Cardiometabolic Effects of Testosterone in Transmen and Estrogen Plus Cyproterone Acetate in Transwomen. The Journal of Clinical Endocrinology and Metabolism 104(6): 1937–1947	Excluded on population – adult study, participants not 18 years or less.
White Hughto JM and Reisner SL (2016) A Systematic Review of the Effects of Hormone Therapy on Psychological Functioning and Quality of Life in Transgender Individuals. Transgender Health 1(1): 21–31	Exclude on population – all included studies conducted in adult population.
Wiepjes CM, de Blok CJM, Staphorsius AS et al. (2020) Fracture Risk in Trans Women and Trans Men Using Long-Term Gender-Affirming Hormonal Treatment: A Nationwide Cohort Study. Journal of Bone and Mineral Research 35(1): 64–70	Excluded on population – adult study, all participants started gender-affirming hormones after 18 years.
Wierckx K, Mueller S, Weyers S et al. (2012) Long-term evaluation of cross-sex hormone treatment in	Excluded on population – adult study, participants not 18 years or less.

Study reference	Reason for exclusion
transsexual persons. The Journal of Sexual Medicine 9(10): 2641–51	
Wierckx K, Van Caenegem E, Schreiner T et al. (2014) Cross-sex hormone therapy in trans persons is safe and effective at short-time follow-up: results from the European network for the investigation of gender incongruence. The journal of sexual medicine 11(8): 1999–2011	Excluded on population – adult study, participants not 18 years or less.
Wilson R, Jenkins C, Miller H et al. (2006) The effect of oestrogen on cytokine and antioxidant levels in male to female transsexual patients. Maturitas 55(1): 14–8	Excluded on population – adult study, participants not 18 years or less.
Witcomb GL, Bouman WP, Claes L et al. (2018) Levels of depression in transgender people and its predictors: Results of a large matched control study with transgender people accessing clinical services. Journal of Affective Disorders 235: 308–315	Excluded on population – although study included some younger people (age range 15 to 79), most participants were adults (mean age around 30.4 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.

Appendix E Evidence tables

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Achille, C., Taggart, T., Eaton, N.R. et al. (2020) Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: Preliminary results. International Journal of Pediatric Endocrinology 2020(1): 8 **Study location** Single centre, New York, United States **Study type** Prospective longitudinal study **Study aim** To assess the psychological wellbeing and quality of life in children and adolescents who have sought endocrine	Inclusion and exclusion not reported- it appears from the description in the publication that all people referred for gender dysphoria were invited to participate, and the vast majority agreed. Of the 95 treatment naïve people who entered the study, 50 people completed all follow-up questionnaires and were included in the analysis. No description of the 45 people without follow-up data reported. The study included 50 children, adolescents and young adults with gender dysphoria.	**Intervention** Endocrine interventions (the collective term used by authors for puberty suppression and gender-affirming hormones) were introduced as per Endocrine Society and the World Professional Association for	**Critical Outcomes** ***Impact on mental health*** Depression symptoms were assessed using the Center for Epidemiologic Studies Depression Scale (CESD-R). Statistically significant improvements in CESD-R score were observed from baseline (initial assessment; 21.4 points) to about 12 months follow-up (13.9 points; p<0.001). Regression analysis, controlling for reported medicines for mental health problems and engagement in counselling, found no statistically significant change from baseline in transfemales (p=0.27) and transmales (p=0.43). The Patient Health Questionnaire Modified for Teens (PHQ 9_Modified for Teens) was also used to assess depression symptoms. Depression scores improved from baseline (p< 0.001; absolute scores not reported numerically). Regression analysis, controlling for reported medicines for mental health problems and engagement in counselling, found no statistically significant change from baseline in transfemales (p=0.07) and transmales (p=0.67). Suicidal ideation measured using the additional questions from the PHQ 9_Modified for Teens, was presented in 10% (5/50) of participants at baseline and 6% (3/50) at	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record 4. b) no **Domain 2: Comparability** 1. c) no comparator **Domain 3: Outcome** 1. c) self-report 2. a) yes – 6 monthly assessment up to 12 months (preliminary results from an ongoing study) 3. c) Follow up rate less than 80% and no description of those lost **Overall quality is assessed as poor** Other comments: Although regression analysis results for some outcomes were controlled for use of medicines for mental health problems, details of these is not

intervention to help with gender dysphoria. **Study dates** Study recruitment ran from December 2013 to December 2018; study is ongoing	17 transfemales and 33 transmales. Diagnostic criteria for gender dysphoria not reported. Mean age at baseline was 16.2 years (SD 2.2). Mean age at the start of gender-affirming hormone treatment not reported.	Transgender Health (WPATH) guidelines. Puberty suppression was: • GnRH agonist and/or anti-androgens (transfemales) • GnRH agonist or medroxyprogesterone (transmales) Average duration of GnRH analogue treatment not reported. Once eligible, gender-affirming hormones were offered, these were: • Oestradiol (transfemales) • Testosterone (transmales) Doses and route of administration not reported. After about 12-months treatment ('wave 3' in the study): • 24 people (48%) were on gender-affirming hormones alone • 12 people (24%) were on puberty suppression alone	about 12-month follow-up, no statistical analysis reported. The study also reported results by gender: In transfemales, 11.8% (2/17) had suicidal ideation at baseline compared with 5.9% (1/17) at 12-month follow-up (no statistically analysis reported) In transmales, 9.1% (3/33) had suicidal ideation at baseline compared with 6.1% (2/33) at 12-month follow-up (no statistically analysis reported) ***Impact on quality of life*** Quality of Life Enjoyment and Satisfaction Questionnaire (QLES-Q-SF) scores: there was no statistically significant change in score from baseline to about 12-months (p=0.085; absolute scores not reported numerically). Regression analysis, controlling for reported medicines for mental health problems and engagement in counselling, found not statistically significant change from baseline in transfemales (p=0.06) and transmales (p=0.08). *No other critical or important outcomes reported*	reported. Other co-morbidities not reported. Source of funding: None

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
		• 11 people (22%) were on both gender-affirming hormones and puberty suppression • 3 people (6%) were on no endocrine intervention Results not represented separately for the sub-group of people who received gender-affirming hormones. Average duration of treatment with gender-affirming hormones not reported. **Comparison** No comparison group. Change overtime reported.		

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Allen, LR, Watson, LB, Egan, AM et al. (2019) Well-being and suicidality among transgender youth after gender-affirming hormones. Clinical Practice in Pediatric Psychology 7(3): 302-311	The study included adolescents and young adults (age range 13-20 years) who received services for gender dysphoria in a clinic in the United States. Participants were required to have received gender-affirming hormones for	39 participants received gender-affirming hormones only 8 participants received a GnRH analogue followed by gender-affirming hormones. Mean duration of treatment in the gender-	**Critical Outcomes** ***Impact on mental health*** The Ask Suicide-Screening Questions (ASQ) instrument was used to assess suicidality. Following an average of about 12 months treatment with gender-affirming hormones, adjusted mean ASQ score was statistically significantly lower (from 1.11 [standard error	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Study location Single centre, Kansas City, United States **Study type** Retrospective longitudinal study **Study aim** To examine suicidality and general well-being following administration of gender-affirming hormones. **Study dates** Participants first presented to the clinic between 2015 and 2018.	at least 3 months, and have pre-test and final assessment data points. No exclusion criteria reported. In total 47 adolescents and young adults with gender dysphoria were included: 14 transfemales (sex assigned at birth male) and 33 transmales (sex assigned at birth female). Diagnostic criteria for gender dysphoria not reported. Mean age at pre-test (before administration of gender-affirming hormones) was 16.59 years (range 13.73 to 19.04). Mean age at the start of treatment in the sub-group who received gender-affirming hormones-only was 16.72 years. Mean age at the start of treatment with gender-affirming hormones in people who previously received a GnRH	affirming hormones only subgroup was 366 days. Mean duration of gender-affirming hormone treatment in people who had previously received a GnRH analogue was not reported. Mean duration of treatment with a GnRH analogue was not reported. Participants were assessed at the start of treatment and at least 3 months after treatment.	(SE) 0.22] at baseline to 0.27 [SE 0.12] at final assessment; p<0.001). The authors also reported change in ASQ separately for transfemales (from 1.21 [SE 0.36] at baseline to 0.24 [SE 0.19] at final assessment) and transmales (from 1.01 [SE 0.36] at baseline to 0.29 [0.13] at final assessment). There was no statistically significant difference in change from baseline between transfemales and transmales (p=0.79) ***Impact on quality of life*** Assessed using the General Well-Being Scale (GWBS) of the Pediatric Quality of Life Inventory. Following an average of about 12 months treatment with gender-affirming hormones, adjusted mean GWBS score was statistically significantly higher (from 61.7 [SE 2.43] at baseline to 70.23 [2.15] at final assessment; p<0.002). The authors also reported change in GWBS of the Pediatric Quality of Life Inventory for transfemales (from 58.44 [SE 4.09] at baseline to 69.52 [SE 3.62] at final assessment) and transmales (from 64.95 [SE 2.66] at baseline to 70.94 [2.35] at final assessment). There was no statistically significant difference in change from baseline between transfemales and transmales (p=0.32) *No other critical or important outcomes reported*	3. a) secure record 4. b) no **Domain 2: Comparability** 2. c) no comparator **Domain 3: Outcome** 1. b) record linkage 2. a) yes – mean duration of treatment was 366 days 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor** Other comments: None Source of funding: Not reported

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
	analogue was not reported.			

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Kaltiala, R., Heino, E., Tyolajarvi, M. et al. (2020) Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria. Nordic Journal of Psychiatry 74(3): 213-219 **Study location** Single centre, Tampere, Finland **Study type** Retrospective chart review **Study aim** To evaluate the psychosocial functioning and need for mental health treatment during the gender identity diagnostic phase and after about a year on gender-affirming hormones.	The study included adolescents who were referred to the gender identity service before they 18 years old, were diagnosed with gender dysphoria, received gender-affirming hormones and completed a follow-up of approximately 12 months after starting hormones. In total 52 adolescents were included, comprising of 11 transfemales and 41 transmales. Gender dysphoria was diagnosed according to International Classification of Disease 10 (ICD-10). The authors state that the corresponding diagnosis to 'gender dysphoria' in	Intervention referred to as 'hormonal sex reassignment treatment' – details of intervention not reported, although gender-affirming hormones were prescribed to all participants. It is not clear from the study whether additional interventions were prescribed. Medical records reviewed for the 'real-life phase' – the approximately 12 months follow-up period for this population in Finland.	**Critical Outcomes** ***Impact on mental health*** Of the 52 people who received gender-affirming hormones, 50% (26/52) needed mental health treatment before or during the assessment and 46% (24/51) needed mental health treatment during the 12-month 'real life' phase (no statistically significant difference). For specific symptoms / conditions: • depression: 54% (28/52) needed treatment before or during the assessment and 15% (8/52) needed treatment during the 12-month 'real life' phase (statistically significant reduction, p<0.001) • anxiety: 48% (25/52) needed treatment before or during the assessment and 15% (8/52) needed treatment during the 12-month 'real life' phase (statistically significant reduction, p<0.001) • suicidality/self-harm: 35% (18/52) needed treatment before or during the assessment and 4% (2/52) needed treatment during the 12-month 'real life' phase (statistically significant reduction, p<0.001) • conduct problems/antisocial: 14% (7/52) needed treatment before or during the assessment and 6% (3/52) needed treatment during the 12-month 'real life	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. a) yes – 12 month follow-up 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor**

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Study dates 2011 to 2017	the ICD-10 is 'transsexualism'. Mean age at diagnosis 18.1 years (range 15.2 to 19.9)		phase (no statistically significant difference, p= 0.18) • psychotic symptoms/psychosis: 2% (1/52) needed treatment before or during the assessment and 4% (2/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 0.56) • substance abuse: 4% (2/52) needed treatment before or during the assessment and 2% (1/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 0.56) • autism: 12% (6/52) needed treatment before or during the assessment and 6% (3/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 0.30) • ADHD: 10% (5/52) needed treatment before or during the assessment and 2% (1/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 0.09) • eating disorder: 2% (1/52) needed treatment before or during the assessment and 2% (1/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 1.0). No details of actual treatment reported. **Important Outcomes** ***Psychosocial Impact*** Study reported on measures of functioning in different domains of adolescent development, reported over the approximately 12-month period after starting gender-affirming	Other comments: None Source of funding: No source of funding reported

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			hormones (referred to as the 'real-life phase' in Finland) Significantly fewer participants were living with parent(s)/ guardians during the real-life phase (40%; 21/50) compared with during gender identity assessment (73%; 38/52; p=0.001)) There was a statistically significant reduction in the number of participants with normative peer contacts, from gender identity assessment (89%; 46/52) to the real-life phase (81%; 42/52; p<0.001). There was no significant difference in the number of participants who were progressing normally in school or work during gender identity assessment (64%; 33/52) compared with the real-life phase (60%; 31/52). There was no significant difference in the number of participants who have been dating or were in steady relationships during gender identity assessment (62%; 32/50) compared with the real-life phase (58%; 30/52). There was no significant difference in the number of participants who were able to deal with matters outside of the home in an age-appropriate manner during gender identity assessment (81% (42/52) compared with the real-life phase (81%; 42/52) *No other critical or important outcomes reported*	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Khatchadourian K, Amed S, Metzger DL (2014) Clinical management of youth with gender dysphoria in Vancouver. The Journal of pediatrics 164(4): 906-11 **Study location** Single centre study, Vancouver, Canada **Study type** Retrospective chart review **Study aim** To describe the patient characteristics, clinical management, and response to treatment in a cohort of people seen in a single clinic. **Study dates** 1998 to 2011	Inclusion criteria were at least Tanner stage 2 pubertal development, previous assessment by a mental health professional and a confirmed diagnosis of gender dysphoria (diagnostic criteria not specified). No exclusion criteria are specified. 63 children, adolescents and young people with gender dysphoria who started gender-affirming hormones, out of 84 young people seen in the unit between 1998 and 2011. 39 transfemales and 24 transmales. Diagnostic criteria for gender dysphoria not reported. Mean age at the start of gender-affirming hormone treatment was 17.4 years (SD 1.9).	**Intervention** Transfemales: Oestrogen (oral micronized 17β-oestradiol) Transmales: Testosterone (injectable testosterone enanthate and/or cypionate) 19 participants (30%) had previously received a GnRH analogue. The median time from start of GnRH analogue to start of gender-affirming hormones was 11.3 months (range 2.2 to 42.0). 11 participants continued GnRH analogues after starting gender-affirming hormones. Average duration of treatment with a GnRH analogue not reported **Comparison** No comparator	**Critical Outcomes** No critical outcomes assessed. **Important outcomes** ***Safety*** Of the 63 participants who received gender-affirming hormones: • No participants permanently discontinued gender-affirming hormones • 3 participants (5%) temporarily discontinued treatment: ○ 2 transmales due to concomitant mental health comorbidities ○ 1 transmale due to androgenic alopecia. ○ No transfemale stopped treatment. The authors report that all patients eventually restarted gender-affirming hormones, although they do not report how long treatment was	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. b) no – although follow-up time is reported for patients with more than 1 clinic visit, duration of treatment with gender-affirming hormones is not reported 3. c) incomplete - missing data **Overall quality is assessed as poor** Other comments: Mental health comorbidity was reported for all participants but not for the gender-affirming hormone cohort separately.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			stopped for, or what the effect of stopped treatment was. • No participants reported major complications • 12 participants (19%) had minor complications: ○ 7 transmales had severe acne (requiring isotretinoin) ○ 1 transmale had andogenic alopecia ○ 3 transmales had mild dyslipidaemia (levels not reported) ○ 1 transmale had significant mood swings ○ No transfemales had minor complications	Concomitant use of other medicines was not reported. Source of funding: No source of funding identified.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Klaver, Maartje, de Mutsert, Renee, van der Loos, Maria A T C et al. (2020) Hormonal Treatment and Cardiovascular Risk Profile in Transgender Adolescents. Pediatrics 145(3) **Study location** Single centre, Amsterdam, Netherlands **Study type**	Participants were included if i) they had started GnRH analogue treatment before 18 years, ii) if whole body dual-energy radiograph absorptiometry was performed at least once during treatment (4 months before or after the start of GnRH analogues or gender-affirming hormones, or within 1.5 years before or after the 22nd birthday), iii) if	Transfemales: Oestrogen (17-β oestradiol [E2]) orally, starting with 5 mcg/kg body weight per day, which was increased every 6 months until the maintenance dose of 2 mg per day was reached. Transmales: mixed testosterone esters (Sustanon), 25 mg/m^2 body surface area every 2 weeks intramuscularly,	**Critical Outcomes** No critical outcomes assessed. **Important outcomes** ***Safety*** Safety outcomes reported separately for transfemales and transmales. **For transfemales**, from the start of gender-affirming hormone treatment to age 22 years: • Mean BMI statistically significantly increased (mean change +1.9, 95% CI 0.6 to 3.2, p<0.005; mean BMI at 22 years= 23.2, 95% CI 21.6 to 24.8). At age 22 years, 9.9% of the cohort were	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Retrospective chart review **Study aim** To examine the effects of treatment on changes in cardiovascular risk factors, including BMI, blood pressure, insulin sensitivity, and lipid levels. **Study dates** 1998-2015	they were likely to have had at least 1 medical consultation in young adulthood. The study included 192 young people with dysphoria who met the above inclusion criteria: 71 transfemales and 121 transmales. Gender dysphoria was diagnosed according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria. Mean age at the start of gender-affirming hormones was 16.4 years (SD 1.1) for transfemales and 16.9 years (SD 0.9) for transmales.	increased every 6 months to maintenance dose of 250 mg every 3 to 4 weeks. When GnRH analogues were started after the age of 16 years a different hormone starter dose was used (1 mg oestrogen daily and 75 mg testosterone weekly). Median (IQR) duration of GnRH analogue (monotherapy) was 2.1 years (1.0 to 2.7) in transfemales and 1.0 (0.5 to 2.9) for transmales.	obese, compared with 3.0% in reference cisgender population[1]. • Mean systolic blood pressure (SBP) did not significantly change (mean change -3 mmHg, 95% CI -8 to 2; mean SBP at 22 years= 117 mmHg, 95% CI 113 to 122) • Mean diastolic blood pressure (DBP) statistically significantly increased (mean change +6 mmHg, 95% CI 3 to 10, p<0.001; mean DBP at 22 years= 75 mmHg, 95% CI 72 to 78) • Mean glucose level did not significantly change (mean change +0.1 mmol/L, 95% CI -0.1 to 0.2; mean glucose level at 22 years= 5.0 mmol/L, 95% CI 4.8 to 5.1) • Mean insulin level did not significantly change (mean change +2.7 mU/L, 95% CI -1.7 to 7.1; mean insulin level at 22 years= 5.0 mU/L (4.8 to 5.1) • Insulin resistance (mean Homeostatic Model Assessment of Insulin Resistance [HOMA-IR]) did not significantly change (mean change +0.7, 95% CI -0.2 to 1.5; mean HOMA-IR at 22 years 2.9, 95% CI 1.9 to 3.9) • Mean total cholesterol did not significantly change (mean change +0.1 mmol/L, 95% CI -0.2 to 0.4; mean total cholesterol at 22 years 4.1 mmol/L, 95% CI 3.8 to 4.4) • Mean HDL cholesterol did not significantly change (mean change +0.0 mmol/L, 95% CI -0.1 to 0.2; mean HDL cholesterol at 22 years 1.6 mmol/L, 95% CI 1.4 to 1.7) • Mean LDL cholesterol did not significantly change (mean change +0.0 mmol/L, 95%	of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. a) yes- follow-up from start of gender-affirming hormones to age 22 years, around 5 years 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor** Other comments: None Source of funding: No external funding

			CI -0.3 to 0.2; mean LDL cholesterol at 22 years 2.0 mmol/L, 95% CI 1.8 to 2.3) • Mean triglycerides statistically significantly increased (mean change +0.2 mmol/L, 95% CI 0.0 to 0.5, p<0.05; triglyceride level at 22 years 1.1 mmol/L, 95% CI 0.9 to 1.4) **For transmales**, from the start of gender-affirming hormone treatment to age 22 years: • Mean BMI statistically significantly increased (mean change +1.4, 95% CI 0.8 to 2.0, p<0.005; mean BMI at 22 years= 23.9, 95% CI 23.0 to 24.7). At age 22 years, 6.6% of the cohort were obese, compared with 2.2% in reference cisgender population[1]. • Mean systolic blood pressure (SBP) statistically significantly increased (mean change +5 mmHg, 95% CI 1 to 9; mean SBP at 22 years= 126 mmHg, 95% CI 122 to 130) • Mean diastolic blood pressure (DBP) statistically significantly increased (mean change +6 mmHg, 95% CI 4 to 9, p<0.001; mean DBP at 22 years= 74 mmHg, 95% CI 72 to 77) • Mean glucose level did not significantly change (mean change 0.0 mmol/L, 95% CI -0.2 to 0.2; mean glucose level at 22 years= 4.8 mmol/L, 95% CI 4.7 to 5.0) • Mean insulin level statistically significantly decreased (mean change -2.1 mU/L, 95% CI -3.9 to -0.3, p<0.05; mean insulin level at 22 years= 8.6 mU/L (6.9 to 10.2) • Insulin resistance (mean Homeostatic Model Assessment of Insulin Resistance [HOMA-IR]) statistically significantly decreased (mean change -0.5, 95% CI -	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			1.0 to -0.1, p<0.05; mean HOMA-IR at 22 years 1.8, 95% CI 1.4 to 2.2) • Mean total cholesterol statistically significantly increased (mean change +0.4 mmol/L, 95% CI 0.2 to 0.6, p<0.001; mean total cholesterol at 22 years 4.6 mmol/L, 95% CI 4.3 to 4.8) • Mean HDL cholesterol statistically significantly decreased (mean change -0.3 mmol/L, 95% CI -0.4 to -0.2, p<0.001; mean HDL cholesterol at 22 years 1.3 mmol/L, 95% CI 1.2 to 1.3) • Mean LDL cholesterol statistically significantly increased (mean change +0.4 mmol/L, 95% CI 0.2 to 0.6, p<0.001; mean LDL cholesterol at 22 years 2.6 mmol/L, 95% CI 2.4 to 2.8) • Mean triglycerides statistically significantly increased (mean change +0.5 mmol/L, 95% CI 0.3 to 0.7, p<0.001; triglyceride level at 22 years 1.3 mmol/L, 95% CI 1.1 to 1.5)	

[1] Reference population taken from Fredriks et al. (2000)

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Klink D, Caris M, Heijboer A et al. (2015) Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. The Journal of Clinical Endocrinology and Metabolism 100(2): e270-5 **Study location** Single centre, Amsterdam, Netherlands **Study type** Retrospective longitudinal study **Study aim** To assess peak bone mass in young adults with gender dysphoria who had received GnRH analogues and gender-affirming hormones during their pubertal years. **Study dates**	34 young people with gender dysphoria who received GnRH analogues, gender-affirming hormones and gonadectomy. The study included 15 transfemales and 19 transmales; mean age at start of gender-affirming hormones was 16.6 years (SD 1.4) and 16.4 years (SD 2.3) respectively. Participants were required to meet the DSM-IV-TR criteria for gender identity disorder of adolescence. Participants were included if they had undergone gonadectomy between June 1998 and August 2012, and they were at least 21 years old when they had the surgery. Bone mineral density data were also required at the start of GnRH analogue, gender-affirming hormones and at the age of 22 years. No concomitant treatments were reported.	**Intervention** Transfemales - oral 17-β oestradiol (incremental dosing) Transmales – IM testosterone (Sustanon 250 mg/ml; incremental dosing) Median duration of treatment with gender-affirming hormones for transfemales was 5.8 years (range 3.0 to 8.0) and for transmales was 5.4 years (range 2.8 to 7.8). The GnRH analogue was SC triptorelin 3.75 mg every 4 weeks. No details of gonadectomy reported. **Comparison** No comparison group. Comparison over time reported.	**Critical outcomes** No critical outcomes reported **Important outcomes** **Safety** ***Bone density: lumbar spine*** **Lumbar spine bone mineral apparent density (BMAD)** Change from starting gender-affirming hormones to age 22 years in transfemales- Mean (SD); g/m^3 • Start of gender-affirming hormones: 0.22 (0.02) • Age 22 years: 0.23 (0.03) • p=0.003 z-score (range) • Start of gender-affirming hormones: -0.90 (0.80) • Age 22 years: -0.78 (1.03) • No statistically significant difference Change from starting gender-affirming hormones to age 22 years in transmales- Mean (SD); g/m^3 • Start of gender-affirming hormones: 0.24 (0.02) • Age 22 years: 0.25 (0.28 • p=0.001 z-score (SD) • Start of gender-affirming hormones: -0.50 (0.81) • Age 22 years: -0.033 (0.95) • p=0.002	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. a) yes – mean duration of gender-affirming hormone treatment was 5.8 and 5.4 years. 3. c) follow-up rate variable across timepoints and no description of those lost **Overall quality is assessed as poor** Other comments: Within person comparison. Small numbers of participants in each subgroup. No

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Gonadectomy took place between June 1998 and August 2012	At the start of gender-affirming hormone treatment, in the transfemale subgroup the median Tanner P was 4 (IQR 2) and the median Tanner G was 12 (IQR 11). In the transmale subgroup the median Tanner B was 5 (IQR 2) and the median Tanner P was 5 (IQR 0).		**Lumbar spine bone mineral density (BMD)** Change from starting gender-affirming hormones to age 22 years in transfemales-Mean (SD); g/m^2 • Start of gender-affirming hormones: 0.84 (0.11) • Age 22 years: 0.93 (0.10) • p<0.001 z-score (range) • Start of gender-affirming hormones: -1.01 (0.98) • Age 22 years: -1.36 (0.83) • No statistically significant difference Change from starting gender-affirming hormones to age 22 years in transmales-Mean (SD); g/m^2 • Start of gender-affirming hormones: 0.91 (0.10) • Age 22 years: 0.99 (0.13) • P<0.001 z-score (range) • Start of gender-affirming hormones: -0.72 (0.99) • Age 22 years: -0.33 (1.12) • No statistically significant difference ***Bone density: femoral region, nondominant side*** **Femoral region, nondominant side BMAD** Change from starting gender-affirming hormones to age 22 years in transfemales-Mean (SD); g/m^3 • Start of gender-affirming hormones: 0.26 (0.04) • Age 22 years: 0.28 (0.05) • No statistically significant difference z-score (SD)	concomitant treatments or comorbidities were reported. Source of funding: None disclosed

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			• Start of gender-affirming hormones: -1.57 (1.74) • Age 22 years: Not reported • No statistical analysis reported Change from starting gender-affirming hormones to age 22 years in transmales-Mean (SD); g/m^3 • Start of gender-affirming hormones: 0.31 (0.04) • Age 22 years: 0.33 (0.05) • p=0.010 z-score (SD) • Start of gender-affirming hormones: -0.28 (0.74) • Age 22 years: Not reported • No statistical analysis reported **Femoral region, nondominant side BMD** Change from starting gender-affirming hormones to age 22 years in transfemales-Mean (SD); g/m^2 • Start of gender-affirming hormones: 0.87 (0.08) • Age 22 years: 0.94 (0.11) • P=0.009 z-score (SD) • Start of gender-affirming hormones: -0.95 (0.63) • Age 22 years: -0.69 (0.74) • No statistically significant difference Change from starting gender-affirming hormones to age 22 years in transmales-Mean (SD); g/m^2 • Start of gender-affirming hormones: 0.88 (0.09) • Age 22 years: 0.95 (0.10) • P<0.001 z-score (SD)	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			• Start of gender-affirming hormones: -0.35 (0.79) • Age 22 years: -0.35 (0.74) • p=0.006	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Kuper, Laura E, Stewart, Sunita, Preston, Stephanie et al. (2020) Body Dissatisfaction and Mental Health Outcomes of Youth on Gender-Affirming Hormone Therapy. Pediatrics 145(4) **Study location** Single centre, Texas, USA **Study type** Prospective longitudinal study **Study aim** To: • explore how baseline body dissatisfaction, depression, and anxiety symptoms vary by gender, age at initial assessment, and	148 children and adolescents with gender dysphoria, n=148, of whom: • 25 received puberty suppression only • 93 received gender-affirming hormone therapy only • 30 received both Results for treatments reported separately. Mean age at initial assessment was 15.4 years (range 9 to 18). Mean age at start of gender-affirming hormone therapy was 16.2 years (range 13.2 to 18.6). All participants met the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria for gender	Hormone therapy, guided by Endocrine Society Clinical Practice Guidelines Follow-up at least 18 months from initial assessment at the clinic. Mean duration of gender-affirming hormone therapy before follow-up was 10.9 months (range 1 to 18; SD 3.3)	**Critical Outcomes** ***Impact on mental health*** Mean depression score, assessed using the Quick Inventory of Depressive Symptoms (QIDS), self-reported was 9.6 (SD 5.0) at baseline and 7.4 (SD 4.5) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean depression score, assessed using the QIDS, clinician-reported was 5.9 (SD 4.1) at baseline and 6.0 (SD 3.8) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean anxiety score, assessed using the Screen for Child Anxiety Related Emotional Disorders (SCARED) questionnaire was 32.6 (SD 16.3) at baseline and 28.4 (SD 15.9) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. d) assessors not blinded to treatment 2. a) yes – follow-up at least 18 months from initial assessment. Mean duration of gender-affirming hormone

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Tanner stage at first medical visit • examine how body dissatisfaction, depression, and anxiety symptoms change over the first year of gender-affirming hormone treatment • explore how any changes vary by affirmed gender, Tanner stage, age, type of treatment, months on gender-affirming hormone therapy, mental health treatment received, and whether chest surgery was also obtained (among transmales). **Study dates** Initial participant assessments took place between August 2014 and March 2018.	dysphoria. Specific inclusion and exclusion criteria for the study are not reported. It would appear that all children and adolescents eligible for gender-affirming hormones were considered eligible for the study. The authors state that before initial assessment with a psychologist, psychiatrist, and/or clinical therapist, parents completed a phone intake survey. Around one-third of families did not follow-up after the phone intake.		hormones and it is unclear whether the change in score was statistically significant. Mean panic score, assessed using specific questions from the SCARED questionnaire was 8.1 (SD 6.3) at baseline and 7.1 (SD 6.5) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean generalised anxiety score, assessed using specific questions from the SCARED questionnaire was 10.0 (SD 5.1) at baseline and 8.8 (SD 6.5) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean social anxiety score, assessed using specific questions from the SCARED questionnaire was 8.5 (SD 4.1) at baseline and 7.7 (SD 4.2) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean separation anxiety score, assessed using specific questions from the SCARED questionnaire was 3.5 (SD 3.0) at baseline and 3.1 (SD 2.5) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-	treatment was 10.9 months. 3. c) patient numbers vary by outcome with no explanation **Overall quality is assessed as poor** Other comments: None Source of funding: Supported by Children's Health. The Research Electronic Data Capture database was funded by the Clinical and Translational Science Awards program

			affirming hormones and it is unclear whether the change in score was statistically significant. Mean school avoidance score, assessed using specific questions from the SCARED questionnaire was 2.6 (SD 2.1) at baseline and 2.0 (SD 2.0) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. The authors also reported results separately for transfemales and transmales: **Transfemales** No statistical analyses were reported for this sub-group and it is unclear whether any changes in score were statistically significant. • Mean depression symptoms, assessed using the QIDS, self-reported was 7.5 (SD 4.9) at baseline and 6.6 (SD 4.4) at follow-up. • Mean depression symptoms, assessed using the QIDS, clinician-reported was 4.2 (SD 3.2) at baseline and 5.4 (SD 3.4) at follow-up. • Mean anxiety symptoms, assessed using the SCARED questionnaire was 26.4 (SD 14.2) at baseline and 24.3 (SD 15.4) at follow-up. • Mean panic symptoms, assessed using specific questions from the SCARED questionnaire was 5.7 (SD 4.9) at baseline and 5.1 (SD 4.9) at follow-up. • Mean generalised anxiety symptoms, assessed using specific questions from	

			the SCARED questionnaire was 8.6 (SD 5.1) at baseline and 8.0 (SD 5.1) at follow-up. • Mean social anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 7.1 (SD 3.9) at baseline and 6.8 (SD 4.4) at follow-up. • Mean separation anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 3.4 (SD 3.3) at baseline and 2.7 (SD 2.3) at follow-up. • Mean school avoidance symptoms, assessed using specific questions from the SCARED questionnaire was 1.8 (SD 1.7) at baseline and 1.9 (SD 2.1) at follow-up. **Transmales** No statistical analyses were reported for this sub-group and it is unclear whether any changes in score were statistically significant. • Mean depression symptoms, assessed using the QIDS, self-reported was 10.4 (SD 5.0) at baseline and 7.5 (SD 4.5) at follow-up. • Mean depression symptoms, assessed using the QIDS, clinician-reported was 6.7 (SD 4.4) at baseline and 6.2 (SD 4.1) at follow-up. • Mean anxiety symptoms, assessed using the SCARED questionnaire was 35.4 (SD 16.5) at baseline and 29.8 (SD 15.5) at follow-up. • Mean panic symptoms, assessed using specific questions from the SCARED questionnaire was 9.3 (SD 6.5) at baseline and 7.9 (SD 6.5) at follow-up.	

			• Mean generalised anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 10.4 (SD 5.0) at baseline and 9.0 (SD 5.1) at follow-up. • Mean social anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 8.5 (SD 4.0) at baseline and 7.8 (SD 4.1) at follow-up. • Mean separation anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 4.2 (SD 3.4) at baseline and 3.4 (SD 2.6) at follow-up. • Mean school avoidance symptoms, assessed using specific questions from the SCARED questionnaire was 2.6 (SD 2.1) at baseline and 2.0 (SD 2.0) at follow-up. No difference in impact on mental health found by Tanner age. Numerical results, statistical analysis and information on specific outcomes not reported. It is unclear from the paper whether Tanner age is at initial assessment, start of GnRH analogues, start of gender-affirming hormones, or another timepoint. **Important Outcomes** ***Impact on body image*** Mean Body Image Scale (BIS) score was 70.7 (SD 15.2) at baseline and 51.4 (SD 18.3) at follow-up. The authors do not present statistical analysis for this population and it is unclear whether the change in score was statistically significant.	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			The authors also reported body image results separately for transfemales and transmales. No statistical analyses were reported for this sub-groups and it is unclear whether changes in score were statistically significant. • In transfemales, BIS score was 67.5 (SD 19.5) at baseline and 49.0 (SD 21.6) at follow-up. • In transmales, BIS score was 71.1 (SD 13.4) at baseline and 52.9 (SD 16.8) at follow-up. No difference in body image score found by Tanner age. Numerical results, statistical analysis and information on specific outcomes not reported. It is unclear from the paper whether Tanner age is at initial assessment, start of GnRH analogues, start of gender-affirming hormones, or another timepoint. *No other critical or important outcomes reported*	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Study dates Lopez de Lara, D., Perez Rodriguez, O., Cuellar Flores, I. et al. (2020) Psychosocial assessment in transgender adolescents. Anales de Pediatria	23 adolescents with gender dysphoria; 16 transmale and 7 transfemale. Participants were required to be at a stage of pubertal development of Tanner 2 or higher. People with mental	Gender-affirming hormones- • Oral oestradiol • Intramuscular testosterone Participants had previously received gonadotropin-releasing	**Critical Outcomes** ***Impact on gender dysphoria*** Following gender-affirming hormones for 12 months, mean (±SD) Utrecht Gender Dysphoria Scale (UGDS) score statistically	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. Not applicable – although a control group is reported

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Study location Single centre in Madrid, Spain **Study type** Prospective analytical study **Study aim** To assess the psychosocial status of patients seeking care in the paediatric endocrinology clinic for gender dysphoria, and the impact on psychosocial status of gender-affirming hormone therapy at 12 months of treatment **Study dates** Not reported	health comorbidity that could affect the experience of gender dysphoria were excluded. Mean age at baseline was 16 years (range 14 to 18). 30 cisgender controls, matched for age, ethnicity, and socioeconomic status	hormone (GnRH) analogues in the intermediate pubertal stages (Tanner 2—3).	significantly improved, from 57.1 (±4.1) at baseline to 14.7 (±3.2; p<0.001) ***Impact on mental health*** Mean depression score statistically significantly improved following treatment with gender-affirming hormones. Mean Beck Depression Inventory II (BDI-II) score (±SD) reduced from 19.3 points (±5.5) at baseline to 9.7 points (±3.9) at 12 months (p<0.001). Mean anxiety scores statistically significantly improved following treatment with gender-affirming hormones. Mean (±SD) State-Trait Anxiety Inventory (STAI) State subscale score improved from 33.3 points (±9.1) at baseline to 16.8 points (±8.1) at 12 months (p<0.001). Mean (±SD) State-Trait Anxiety Inventory (STAI) Trait subscale score improved from 33.0 points (±7.2) at baseline to 18.5 points (±8.4) at 12 months (p<0.001). **Important Outcomes** ***Psychosocial Impact*** There was not change in family functioning, measured using the Family APGAR test, from baseline (17.9 points) to 1 year after starting gender-affirming hormones (18.0 points; no statistical analysis reported). Results from the Strengths and Difficulties Questionnaire, Spanish Version (SDQ-Cas) showed statistically significant improvements from baseline (14.7 points; SD±3.3) to 12	on, people in this group did not have gender dysphoria. 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. Not applicable – although a control group is reported on, people in this group did not have gender dysphoria. **Domain 3: Outcome** 1. d) assessors not blinded to treatment 2. a) yes – 12 months treatment with gender-affirming hormones 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor** Other comments: None Source of funding: Not reported

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			months after gender-affirming hormones (10.3 points; SD±2.9; p<0.001) *No other critical or important outcomes reported*	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Stoffers, Iris E; de Vries, Martine C; Hannema, Sabine E (2019) Physical changes, laboratory parameters, and bone mineral density during testosterone treatment in adolescents with gender dysphoria. The journal of sexual medicine 16(9): 1459-1468 **Study location** Single centre, Leiden, Netherlands **Study type** Retrospective chart review **Study aim** To report changes in height, BMI, blood pressure, laboratory parameters and bone density. **Study dates** November 2010 to August 2018	62 transmales with gender dysphoria. participants were required to have been receiving testosterone therapy for at least 6 months. Further inclusion or exclusion criteria not reported. Gender dysphoria was diagnosed according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria.	Testosterone intramuscular injection (Sustanon 250 mg). Dose escalated every 6 months up to the standard adult dose of 125 mg every 2 weeks or 250 mg every 3-4 weeks. A more rapid dose escalation was using in patients who started GnRH analogue treatment at 16 years or older. Median age at start of testosterone treatment was 17.2 years (range 14.9 to 18.4) Median duration of testosterone treatment was 12 months (range 5 to 33) Median duration of GnRH analogue treatment was 8 months (range 3 to 39)	**Critical Outcomes** No critical outcomes assessed. **Important outcomes** ***Safety*** ***Bone mineral density (BMD): lumbar spine*** There was no statistically significant difference in lumber spine bone mineral density (BMD) from start of testosterone treatment to any timepoint, up to 24 months follow-up. Mean (±SD), g/cm^2: • Start of testosterone: 0.90 (±0.11) • 6 months: 0.94 (±0.10) • 12 months: 0.95 (±0.09) • 24 months: 0.95 (±0.11) z-score (±SD): • Start of testosterone: -0.81 (±1.02) • 6 months: -0.67 (±0.95) • 12 months: -0.66 (±0.81) • 24 months: -0.74 (±1.17) ***Bone mineral density (BMD): femoral neck (hip)*** There was no statistically significant difference in right or left femoral neck (hip) bone mineral density (BMD) from start of	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. a) yes – mean duration of gender-affirming hormone treatment was 5.8 and 5.4 years. 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor** Other comments: None Source of funding: None

			testosterone treatment to any timepoint, up to 24 months follow-up. **Right** Mean (±SD), g/cm^2: • Start of testosterone: 0.77 (±0.08) • 6 months: 0.84 (±0.11) • 12 months: 0.82 (±0.08) • 24 months: 0.85 (±0.11) z-score (±SD): • Start of testosterone: -0.97 (0.79) • 6 months: -0.54 (±0.96) • 12 months: -0.80 (±0.69) • 24 months: -0.31 (±0.84) **Left** Mean (±SD), g/cm^2: • Start of testosterone: 0.76 (±0.09) • 6 months: 0.83 (±0.12) • 12 months: 0.81 (±0.08) • 24 months: 0.86 (±0.09) z-score (±SD): • Start of testosterone: -1.07 (0.85) • 6 months: -0.62 (±1.12) • 12 months: -0.93 (±0.63) • 24 months: -0.20 (±0.70) ***Other safety-related outcomes*** • Alkaline phosphatase: statistically significant increases observed from start of testosterone treatment to 6 months and 12 months ($p<0.001$), although difference at 24 months was not statistically significant, Median (IQR), U/L ○ Start of testosterone: 102 (78 to 136) ○ 6 months: 115 (102 to 147) ○ 12 months: 112 (88 to 143) ○ 24 months: 81 (range 69 to 98) • Creatinine: statistically significant increases observed from start of testosterone treatment to 6, 12 and	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			24 months (p<0.001). Mean (±SD), umol/L ○ Start of testosterone: 62 (±7) ○ 6 months: 70 (±9) ○ 12 months: 74 (±10) ○ 24 months: 81 (±10) There was no statistically significant change from start of testosterone treatment in: • HbA1c • Aspartate aminotransferase (AST) • Alanine aminotransferase (ALT) • Gamma-glutamyl transferase • Urea Numerical results, follow-up duration and further details of statistical analysis not reported.	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Vlot MC, Klink DT, den Heijer M et al. (2017) Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone 95: 11-19 **Study location** Single centre, Amsterdam, Netherlands **Study type** Retrospective chart review **Study aim** To investigate the impact of GnRH analogues and gender-affirming hormones on bone mineral apparent density (BMAD) in transgender adolescents. The study also report on levels of bone turnover markers, although the authors concluded that the	70 adolescents with gender dysphoria (42 transmales and 28 transfemales). Median age (range) at the start of gender-affirming hormones was 16.3 years (15.9 to 19.5) for transmales and 16.0 years (14.0 to 18.9) for transfemales. Participants were included if they had a diagnosis of gender dysphoria according to DSM-IV-TR criteria who received GnRH analogues and then gender-affirming hormones. No concomitant treatments were reported. The study categorised participants into a young and old pubertal group, based on their bone age. The young transmales had a bone age of <14 years and the old transmales had a bone age of ≥14 years. The young transfemales group had a bone age of	Transfemales: Oestradiol oral Dose escalated every 6 months until standard adult dose of 2 mg daily was reached Transmales: Testosterone intramuscular injection (Sustanon 250 mg). Dose escalated every 6 months up to the standard adult dose of 250 mg every 4 weeks or 250 mg every 3-4 weeks. All participants previously received a GnRH analogue (triptorelin 3.75 mg subcutaneously every 4 weeks) Median duration of GnRH analogue therapy not reported.	**Critical outcomes** No critical outcomes reported **Important outcomes** ***Bone density: lumbar spine*** **Lumbar spine bone mineral apparent density (BMAD)** Transfemales (bone age <15 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones (C0): 0.20 (0.18 to 0.24) • 24-month follow-up (C24): 0.22 (0.19 to 0.27) • Statistically significant increase (p≤0.01) z-score (range) • Start of gender-affirming hormones (C0): -1.52 (-2.36 to 0.42) • 24-month follow-up (C24): • Statistically significant increase (p≤0.05) Transfemales (bone age ≥15 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones: 0.22 (0.19 to 0.24) • 24-months: 0.23 (0.21 to 0.26) • Statistically significant increase (p≤0.05) z-score (range) • Start of gender-affirming hormones: -1.15 (-2.21 to 0.08) • 24-months: -0.66 (-1.66 to 0.54)	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. a) yes- 24 month follow-up 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor.** Other comments: None Source of funding: grant from Abbott diagnostics

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
added value of these seems to be limited. **Study dates** Participants started gender-affirming therapy between 2001 and 2011	<15 years and the old transfemales group ≥15 years.		Statistically significant increase (p≤0.05) Transmales (bone age <14 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones: 0.23 (0.19 to 0.28) • 24-months: 0.25 (0.22 to 0.28) • Statistically significant increase (p≤0.01) z-score (range) • Start of gender-affirming hormones: -0.84 (-2.2 to 0.87) • 24-months: -0.15 (-1.38 to 0.94) Statistically significant increase (p≤0.01) Transmales (bone age ≥14 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones: 0.24 (0.20 to 0.28) • 24-months: 0.25 (0.21 to 0.30) • Statistically significant increase (p≤0.01) z-score (range) • Start of gender-affirming hormones: -0.29 (-2.28 to 0.90) • 24-months: -0.06 (-1.75 to 1.61) Statistically significant increase (p≤0.01) ***Bone density: femoral neck*** **Femoral neck BMAD** Transfemales (bone age <15 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			• Start of gender-affirming hormones: 0.27 (0.20 to 0.33) • 24-months: 0.27 (0.20 to 0.36) • No statistically significant change z-score (range) • Start of gender-affirming hormones: -1.32 (-3.39 to 0.21) • 24-months: -1.30 (-3.51 to 0.92) • No statistically significant change Transfemales (bone age ≥15 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones: 0.30 (0.26 to 0.34) • 24-months: 0.29 (0.24 to 0.38) • No statistically significant change z-score (range) • Start of gender-affirming hormones: -0.36 (-1.50 to 0.46) • 24-months: -0.56 (-2.17 to 1.29) • No statistically significant change Transmales (bone age <14 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones: 0.30 (0.22 to 0.35) • 24-months: 0.33 (0.23 to 0.37) • Statistically significant increase (p≤0.01) z-score (range) • Start of gender-affirming hormones: -0.37 (-2.28 to 0.47) • 24-months: -0.37 (-2.03 to 0.85)	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			• Statistically significant increase (p≤0.01) Transmales (bone age ≥14 years), change from starting gender-affirming hormones to 24 months follow-up. • Start of gender-affirming hormones: 0.30 (0.23 to 0.41) • 24-months: 0.32 (0.23 to 0.41) • Statistically significant increase (p≤0.01) z-score (range) • Start of gender-affirming hormones: -0.27 ((-1.91 to 1.29) • 24-months: 0.02 (-2.1 to 1.35) • Statistically significant increase (p≤0.05)	

Appendix F Quality appraisal checklists

Newcastle-Ottawa Quality Assessment Form for Cohort Studies

Note: A study can be given a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability.

Selection

1) Representativeness of the exposed cohort
   a) Truly representative (one star)
   b) Somewhat representative (one star)
   c) Selected group
   d) No description of the derivation of the cohort
2) Selection of the non-exposed cohort
   a) Drawn from the same community as the exposed cohort (one star)
   b) Drawn from a different source
   c) No description of the derivation of the non exposed cohort
3) Ascertainment of exposure
   a) Secure record (e.g., surgical record) (one star)
   b) Structured interview (one star)
   c) Written self report
   d) No description
   e) Other
4) Demonstration that outcome of interest was not present at start of study
   a) Yes (one star)
   b) No

Comparability

1) Comparability of cohorts on the basis of the design or analysis controlled for confounders
   a) The study controls for age, sex and marital status (one star)
   b) Study controls for other factors (list) _____ (one star)
   c) Cohorts are not comparable on the basis of the design or analysis controlled for confounders

Outcome

1) Assessment of outcome
   a) Independent blind assessment (one star)
   b) Record linkage (one star)
   c) Self report
   d) No description
   e) Other
2) Was follow-up long enough for outcomes to occur
   a) Yes (one star)
   b) No
   Indicate the median duration of follow-up and a brief rationale for the assessment above:_____
3) Adequacy of follow-up of cohorts
   a) Complete follow up- all subject accounted for (one star)

b) Subjects lost to follow up unlikely to introduce bias- number lost less than or equal to 20% or description of those lost suggested no different from those followed. (one star)
c) Follow up rate less than 80% and no description of those lost
d) No statement

Thresholds for converting the Newcastle-Ottawa scales to AHRQ standards (good, fair, and poor):

Good quality: 3 or 4 stars in selection domain AND 1 or 2 stars in comparability domain AND 2 or 3 stars in outcome/exposure domain

Fair quality: 2 stars in selection domain AND 1 or 2 stars in comparability domain AND 2 or 3 stars in outcome/exposure domain

Poor quality: 0 or 1 star in selection domain OR 0 stars in comparability domain OR 0 or 1 stars in outcome/exposure domain

Appendix G Grade profiles

Table 2: Question 1: For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? - Gender dysphoria

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on gender dysphoria (1 uncontrolled, prospective observational study)									
Change from baseline in mean gender dysphoria score, measured using the UGDS (duration of treatment 12 months). Higher scores indicate greater gender dysphoria.									
1 cohort study Lopez de Lara et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 57.1 (SD 4.1) T1 (12 months) = 14.7 (SD 3.2) Statistically significant improvement, p<0.001	Critical	VERY LOW

Abbreviations: p: p-value; SD: standard deviation; UGDS: Utrecht Gender Dysphoria Scale

1 Downgraded 1 level - the cohort study by Lopez de Lara et al. 2020 was assessed at high risk of bias (poor quality overall; lack of blinding and no control group)

Table 3: Question 1: For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Mental health

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on mental health (3 uncontrolled, prospective observational studies and 2 uncontrolled, retrospective observational studies)									
Change from baseline in mean depression score, measured using the BDI-II (duration of treatment 12 months). Higher scores indicate more severe depression.									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 cohort study Lopez de Lara et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 19.3 (SD 5.5) T1 (12 months) = 9.7 (SD 3.9) Statistically significant improvement, p<0.001	Critical	VERY LOW
Change from baseline in mean depression score, measured using the CESD-R (approximately 12-month follow-up). Higher scores indicate more severe depression.									
1 cohort study Achille et al. 2020	Serious limitations[2]	Serious indirectness[3]	No serious inconsistency	Not calculable	N=50	None	Wave 1 (baseline) = 21.4 Wave 3 (approx. 12 months) = 13.9 Statistically significant improvement (p<0.001)	Critical	VERY LOW
Change from baseline in depression score, measured using the Patient Health Questionnaire Modified for Teens (PHQ 9_Modified for Teens) (approximately 12-month follow-up). Higher scores indicate more severe depression.									
1 cohort study Achille et al. 2020	Serious limitations[2]	Serious indirectness[3]	No serious inconsistency	Not calculable	N=50	None	Statistically significant reductions in mean score, p<0.001 Results presented diagrammatically, numerical results for mean score not reported	Critical	VERY LOW
Change from baseline in depression symptoms, measured using the Quick Inventory of Depressive Symptoms (QIDS), self-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe depression.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=105	None	Baseline = 9.6 (SD 5.0) Follow-up = 7.4 (SD 4.5) No statistical analysis reported for the sub-group of participants receiving gender-affirming hormones	Critical	VERY LOW
Change from baseline in depression symptoms, measured using the Quick Inventory of Depressive Symptoms (QIDS), clinician-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe depression.									
1 cohort study	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=106	None	Baseline = 5.9 (SD 4.1) Follow-up = 6.0 (SD 3.8)	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Kuper et al. 2020							No statistical analysis reported for the sub-group of participants who received gender-affirming hormones		
Need for treatment due to depression, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 54% (28/52) During real life phase 15% (8/52) Statistically significant reduction (p<0.001)	Critical	VERY LOW
Change from baseline in anxiety score, measured using the STAI-State subscale (duration of treatment 12 months). Higher scores indicate more severe anxiety.									
1 cohort study Lopez de Lara et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 33.3 (SD 9.1) T1 (12 months) = 16.8 (SD 8.1) Statistically significant improvement, p<0.001	Critical	VERY LOW
Change from baseline in anxiety score, measured using the STAI-Trait subscale (duration of treatment 12 months). Higher scores indicate more severe anxiety.									
1 cohort study Lopez de Lara et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 33.0 (SD 7.2) T1 (12 months) = 18.5 (SD 8.4) Statistically significant improvement, p<0.001	Critical	VERY LOW
Change from baseline in anxiety symptoms, measured using the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe anxiety.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=80	None	Baseline = 32.6 (SD 16.3) Follow-up = 28.4 (SD 15.9) No statistical analysis reported for the sub-group of participants	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							who received gender-affirming hormones		
Change from baseline in panic symptoms, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=82	None	Baseline = 8.1 (SD 6.3) Follow-up = 7.1 (SD 6.5) No statistical analysis reported for the sub-group of participants who received gender-affirming hormones	Critical	VERY LOW
Change from baseline in generalised anxiety symptoms, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=82	None	Baseline = 10.0 (SD 5.1) Follow-up = 8.8 (SD 5.0) No statistical analysis reported for the sub-group of participants who received gender-affirming hormones	Critical	VERY LOW
Change from baseline in social anxiety symptoms, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=82	None	Baseline = 8.5 (SD 4.1) Follow-up = 7.7 (SD 4.2) No statistical analysis reported for the sub-group of participants who received gender-affirming hormones	Critical	VERY LOW
Change from baseline in separation anxiety symptoms, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=81	None	Baseline = 3.5 (SD 3.0) Follow-up = 3.1 (SD 2.5) No statistical analysis reported for the sub-group of participants	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							who received gender-affirming hormones		
Change from baseline in school avoidance, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=80	None	Baseline = 2.6 (SD 2.1) Follow-up = 2.0 (SD 2.0) No statistical analysis reported for the sub-group of participants who received gender-affirming hormones	Critical	VERY LOW
Need for treatment due to anxiety, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 48% (25/52) During real life phase 15% (8/52) Statistically significant reduction (p<0.001)	Critical	VERY LOW
Change from baseline in adjusted mean suicidality score, measured using the ASQ instrument (mean treatment duration 349 days). Higher scores indicate a greater degree of suicidality.									
1 cohort study Allen et al. 2019	Serious limitations[5]	No serious indirectness	No serious inconsistency	Not calculable	N=39	None	T0 (baseline) = 1.11 (SE 0.22) T1 (final assessment) = 0.27 (SE 0.12) Statistically significant improvement in score from T0 to T1, p<0.001	Critical	VERY LOW
Change from baseline in percentage of participants with suicidal ideation, measured using the additional questions from the PHQ 9_Modified for Teens (approximately 12-month follow-up)									
1 cohort study Achille et al. 2020	Serious limitations[2]	Serious indirectness[3]	No serious inconsistency	Not calculable	N=50	None	Wave 1 (baseline) = 10% (5/50) Wave 3 (approx. 12 months) = 6% (3/50)	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							No statistical analysis reported		
Change from baseline in suicidal ideation (passive), information on which was collected by clinician, exact methods / tools not reported (mean duration of gender-affirming hormone treatment was 10.9 months)									
1 cohort study Kuper et al. 2020	Serious limitations[4]	Serious indirectness [6]	No serious inconsistency	Not calculable	N=130	None	Lifetime = 81% (105 people) 1 month before initial assessment = 25% (33 people) Follow-up period = 38% (51 people) No statistical analysis reported	Critical	VERY LOW
Change from baseline in suicide attempts, information on which was collected by clinician, exact methods / tools not reported (mean duration of gender-affirming hormone treatment was 10.9 months)									
1 cohort study Kuper et al. 2020	Serious limitations[4]	Serious indirectness[6]	No serious inconsistency	Not calculable	N=130	None	Lifetime = 15% (20 people) 3 months before initial assessment = 2% (3 people) Follow-up period = 5% (6 people) No statistical analysis reported	Critical	VERY LOW
Change from baseline in non-suicidal self-injury, information on which was collected by clinician, exact methods / tools not reported (mean duration of gender-affirming hormone treatment was 10.9 months)									
1 cohort study Kuper et al. 2020	Serious limitations[4]	Serious indirectness[6]	No serious inconsistency	Not calculable	N=130	None	Lifetime = 52% (68 people) 3 months before initial assessment = 10% (13 people) Follow-up period = 17% (23 people) No statistical analysis reported	Critical	VERY LOW
Need for treatment due to suicidality / self-harm, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 35% (18/52) During real life phase	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							4% (2/52) Statistically significant reduction (p<0.001)		
Need for mental health treatment, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 50% (26/52) During real life phase 46% (24/51) No statistically significant difference (p= 0.77)	Critical	VERY LOW
Need for treatment due to conduct problems / antisocial, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 14% (7/52) During real life phase 6% (3/52) No statistically significant difference (p= 0.18)	Critical	VERY LOW
Need for treatment due to psychotic symptoms or psychosis, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 2% (1/52) During real life phase 4% (2/52) No statistically significant difference (p= 0.56)	Critical	VERY LOW
Need for treatment due to substance abuse, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 4% (2/52) During real life phase 2% (1/52) No statistically significant difference (p= 0.56)	Critical	VERY LOW
Need for treatment due to autism, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 12% (6/52) During real life phase 6% (3/52) No statistically significant difference (p= 0.30)	Critical	VERY LOW
Need for treatment due to ADHD, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 10% (5/52) During real life phase 2% (1/52) No statistically significant difference (p= 0.09)	Critical	VERY LOW
Need for treatment due to eating disorder, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 2% (1/52)	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							During real life phase 2% (1/52) No statistically significant difference (p=1.0)		

Abbreviations: ADHD: attention deficit hyperactivity disorder; ASQ: Ask Suicide-Screening Questions; CESD-R: Center for Epidemiologic Studies Depression Scale; BDI-II: Beck Depression Inventory II (BDI-II); p: p-value; PHQ 9_Modified for Teens: Patient Health Questionnaire Modified for Teens; SCARED: Screen for Child Anxiety Related Emotional Disorders; SD: standard deviation; STAI: State-Trait Anxiety Inventory

1 Downgraded 1 level - the cohort study by Lopez de Lara et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding and no control group).
2 Downgraded 1 level - the cohort study by Achille et al (2020) was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
3 Serious indirectness in Achille 2020- Outcome reported for full study cohort, of whom 30% were taking no treatment or puberty suppression alone at follow-up. Results for people taking gender-affirming hormones not reported separately.[4] Downgraded 1 level - the cohort study by Kuper et al. (2020) was assessed at high risk of bias (poor quality).
5 Downgraded 1 level - the cohort study by Allen et al. (2019) was assessed at high risk of bias (poor quality; lack of blinding and no control group).
6 Serious indirectness in Kuper et al. 2020- Outcome reported for full study cohort, of whom approximately 17% received puberty suppression alone and did not receive gender-affirming hormones
7 Downgraded 1 level - the cohort study by Kaltiala et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding and no control group).

Table 4: Question 1: For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Quality of life

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on quality of life (1 uncontrolled, prospective observational study and 1 uncontrolled, retrospective observational study)									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Change from baseline in mean quality of life score, measured using the QLES-Q-SF) (approximately 12-month follow-up). Higher scores indicated better quality of life.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=50	None	Numerical improvements in mean score reported from wave 1 (baseline) to wave 3 (approx. 12 months), but difference not statistically significant (p = 0.085) Results presented diagrammatically, numerical results for mean score not reported	Critical	VERY LOW
Change from baseline in adjusted mean well-being score, measured using the GWBS of the Pediatric Quality of Life Inventory (mean treatment duration 349 days). Higher scores indicated better well-being.									
1 cohort study Allen et al. 2019	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=39	None	T0 (baseline) = 61.70 (SE 2.43) T1 (final assessment) = 70.23 (SE 2.15) Statistically significant improvement in well-being score, p<0.002	Critical	VERY LOW

Abbreviations: GWBS: General Well-Being Scale; p: p-value; QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire; SE: standard error

1 Downgraded 1 level - the cohort study by Achille et al (2020) was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
2 Serious indirectness in Achille et al. 2020 - Outcome reported for full study cohort, of whom 30% were taking no treatment or puberty suppression alone at follow-up. Results for people taking gender-affirming hormones not reported separately.
3 Downgraded 1 level - the cohort study by Allen et al. (2019) was assessed at high risk of bias (poor quality; lack of blinding and no control group).

Table 5: Question 1: For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Body image

QUALITY	Summary of findings	IMPORTANCE	CERTAINTY

					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on body image *(1 uncontrolled, prospective observational study)*									
Change from baseline in mean body image, measured using the BIS (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores represent a higher degree of body dissatisfaction.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=86	None	Baseline = 70.7 (SD 15.2) Follow-up = 51.4 (SD 18.3) No statistical analysis reported for the sub-group of participants who received gender-affirming hormones	Important	VERY LOW

Abbreviations: BIS: Body Image Scale; p: p-value; SD: standard deviation

1 Downgraded 1 level - the cohort study by Kuper et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).

Table 6: Question 1: For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Psychological impact

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Psychosocial Impact (1 uncontrolled, prospective observational study and 1 uncontrolled, retrospective observational study)									
Change from baseline in family functioning, measured using the Family APGAR test. Higher scores suggest more family dysfunction.									
1 cohort study Lopez de Lara et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 17.9 T1 (12 months) = 18.0 No statistical analysis reported	Important	VERY LOW
Change from baseline in mean patient strengths and difficulties score, measured using the SDQ, Spanish Version (total difficulties score) (duration of treatment 12 months). Higher scores suggest the presence of a behavioural disorder.									
1 cohort study	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 14.7 (SD 3.3) T1 (12 months) = 10.3 (SD 2.9)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Lopez de Lara et al. 2020							Statistically significant improvement p<0.001		
Functioning in adolescent development: Living with parent(s)/ guardians[2] (outcome reported for the approximately 12-month period after starting gender-affirming hormones; referred to as the 'real-life phase' in Finland). Not living with parent(s) or guardian in your early 20s is a marker of age-appropriate functioning in Finnish culture.									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During gender identity assessment = 73% (38/52) During real life phase = 40% (21/50) Statistically significant reduction (p=0.001)	Important	VERY LOW
Functioning in adolescent development: Normative peer contacts[4] (outcome reported for the approximately 12-month period after starting gender-affirming hormones; referred to as the 'real-life phase' in Finland)									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During gender identity assessment = 89% (46/52) During real life phase = 81% (42/52) Statistically significant reduction (p<0.001)	Important	VERY LOW
Functioning in adolescent development: Progresses normatively in school/ work[5] (outcome reported for the approximately 12-month period after starting gender-affirming hormones; referred to as the 'real-life phase' in Finland)									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During gender identity assessment = 64% (33/52) During real life phase = 60% (31/52) No statistically significant difference (p=0.69)	Important	VERY LOW
Functioning in adolescent development: Has been dating or had steady relationships[6] (outcome reported for the approximately 12-month period after starting gender-affirming hormones; referred to as the 'real-life phase' in Finland)									
1 cohort study	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During gender identity assessment = 62% (32/50)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Kaltiala et al. 2020							During real life phase = 58% (30/52) No statistically significant difference (p=0.51)		
Functioning in adolescent development: Is age-appropriately able to deal with matters outside of the home[7] (outcome reported for the approximately 12-month period after starting gender-affirming hormones; referred to as the 'real-life phase' in Finland)									
1 cohort study Kaltiala et al. 2020	Serious limitations[2]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During gender identity assessment = 81% (42/52) During real life phase = 81% (42/52) No statistically significant difference (p=1.00)	Important	VERY LOW

Abbreviations: APGAR: Adaptability, Partnership, Growth, Affection and Resolve; p: p-value; SD: standard deviation; SDQ: Strengths and Difficulties Questionnaire

1 Downgraded 1 level - the cohort study by Lopez de Lara et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding and no control group).
2 Living arrangements were classified as (1) living with at least one parent/guardian, (2) living in a boarding school, with an adult relative, in some form of supported accommodation or the like, where supervision and guidance by a responsible adult is provided, (3) independently alone or in a shared household with a peer, (4) with a romantic partner. In the analyses dichotomised living arrangements as (a) parent(s)/guardian(s) vs. in other arrangements.
3 Downgraded 1 level - the cohort study by Kaltiala et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding and no control group).
4 Peer relationships were classified as: (1) socialises with friends in leisure time, outside of activities supervised by adults, (2) socialises with peers only at school or in the context of rehabilitative activity, (3) spends time close to peers, for example in school or rehabilitative activity, but does not connect with them, (4) does not meet peers at all. In the analyses, peer relationships during (a) gender identity assessment and (b) the real-life phase were dichotomized to age-appropriate (normative) (1) vs. restricted or lacking (2–4).
5 School/work participation was classified as (1) age appropriate participation in mainstream curriculum, progresses without difficulties, (2) participates in mainstream curriculum with difficulty, (3) participates in rehabilitative educational or work activity, (4) not involved in education and working life. Age-appropriate participation during (1) was recorded if the adolescent attended mainstream secondary education or upper secondary education at a regular rate (a class per year in comprehensive school, has not changed more than once between tracks in upper secondary education) or had proceeded to work life after completing vocational education. Participation with difficulty (2) was recorded if the adolescent was enrolled in mainstream education but had to repeat a class, studied with special arrangements (for example, in a special small group), or followed some form of adjusted curriculum. In the analyses, school/work life during (a) gender identity assessment and (b) real-life phase was dichotomised to normative (1) vs. any other (2, 3 or 4).
6 Romantic involvement was recorded (1) has or has had a dating or steady relationship, not only online, (2) has had a romantic relationship only online, (3) has not had dating or steady relationships. In the analyses we compared has or has had (1) vs. has not had (2,3) a dating or steady relationship during (a) gender identity assessment and (b) real-life phase. Sexual history was recorded in more detail in case histories during gender identity assessment, and for this period we also collected the experiences of (French) kissing (yes/no), intercourse (yes/no) and experience of any genitally intimate contact with a partner (petting under clothes or naked, intercourse, oral sex) (yes/no).

7 In recording age-appropriate competence in managing everyday matters it was expected that early adolescents (up to 14 years) would be able, for example, to do shopping and travel alone on local public transport, and to help with household duties assigned by their parents. Middle adolescents (15–17 years) were further assumed, for example, to be able make telephone calls in matters important to them (for example, when seeking a summer job), to deal with school-related issues with school personnel without parental participation, to select and start new hobbies independently and to fulfil their role in summer jobs and in similar responsibilities of young people. Late adolescents (18 years and over), legally adults, were expected to have, in addition to the above, competence to talk to authorities such as professionals in health and social services, employment or educational institutions, to deal with banks or health insurance, to manage their financial issues and to manage their housekeeping if they chose to move to live independently of parents/guardians. Competence in managing everyday matters was recorded as follows; (1) the adolescent is able to cope age appropriately outside home, (2) the adolescent needs support in age-appropriate matters outside home but functions age-appropriately in the home (manages her/his own hygiene, clothing and nutrition, participates in (younger subjects) or takes responsibility for (older subjects) housekeeping) and (3) the adolescent's functioning is inadequate both at home and outside home. For the analyses, participants were determined to be able to age-appropriately able cope with matters outside of the home (1) vs. not (2,3)

Table 7: Question 2: For children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Bone density

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Lumbar spine bone mineral apparent density (BMAD) (2 uncontrolled, retrospective observational studies)									
Change from start of gender-affirming hormones to age 22 years in lumber spine BMAD in transfemales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=13 (Mean) N=14 (z-score)	None	Mean (SD), g/m^3 Start of gender-affirming hormones: 0.22 (0.02) Age 22 years: 0.23 (0.03) P=0.003 z-score (SD) Start of gender-affirming hormones: -0.90 (0.80) Age 22 years: -0.78 (1.03) No statistically significant difference	Important	VERY LOW
Change from baseline in lumbar spine BMAD in transfemales with a bone age less than 15 years ('young'; 24 months follow-up)									
1 cohort study Vlot et al 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=15	None	Median (range), g/m^3 Start of gender-affirming hormones (C0): 0.20 (0.18 to 0.24)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							24-month follow-up (C24): 0.22 (0.19 to 0.27) Statistically significant increase (p≤0.01) z-score (range) Start of gender-affirming hormones (C0): -1.52 (-2.36 to 0.42) 24-month follow-up (C24): -1.10 (-2.44 to 0.69) Statistically significant increase (p≤0.05)		
Change from baseline in lumbar spine BMAD in transfemales with a bone age of 15 years or more ("old"; 24 months follow-up)									
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=5	None	Median (range), g/m^3 Start of gender-affirming hormones (C0): 0.22 (0.19 to 0.24) 24-month follow-up (C24): 0.23 (0.21 to 0.26) Statistically significant increase (p≤0.05) z-score (range) Start of gender-affirming hormones (C0): -1.15 (-2.21 to 0.08) 24-month follow-up (C24): -0.66 (-1.66 to 0.54) Statistically significant increase (p≤0.05)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in lumber spine BMAD in transmales									

This document was prepared in October 2020

Page 123 of 156

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=19 (Mean and z-score)	None	Mean (SD), g/m³ Start of gender-affirming hormones: 0.24 (0.02) Age 22 years: 0.25 (0.28) P=0.001 z-score Start of gender-affirming hormones: -0.50 (0.81) Age 22 years: -0.033 (0.95) P=0.002	Important	VERY LOW
Change from baseline in lumbar spine BMAD in transmales with a bone age of less than 14 years ('young'; 24 months follow-up)									
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=11	None	Median (range), g/m³ Start of gender-affirming hormones (C0): 0.23 (0.19 to 0.28) 24-month follow-up (C24): 0.25 (0.22 to 0.28) Statistically significant increase (p≤0.01) z-score (range) Start of gender-affirming hormones (C0): -0.84 (-2.2 to 0.87) 24-month follow-up (C24): -0.15 (-1.38 to 0.94) Statistically significant increase (p≤0.01)	Important	VERY LOW
Change from baseline in lumbar spine BMAD in transmales with a bone age of 14 years or more ('old'; 24 months follow-up)									
1 cohort study	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=23	None	Median (range), g/m³	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Vlot et al 2017							Start of gender-affirming hormones (C0): 0.24 (0.20 to 0.28) 24-month follow-up (C24): 0.25 (0.21 to 0.30) Statistically significant increase (p≤0.01) z-score (range) Start of gender-affirming hormones (C0): -0.29 (-2.28 to 0.90) 24-month follow-up (C24): -0.06 (-1.75 to 1.61) Statistically significant increase (p≤0.01)		
Change in femoral neck BMAD (2 uncontrolled, retrospective observational studies)									
Change from start of gender-affirming hormones to age 22 years in femoral neck BMAD in transfemales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=14 (Mean) N=10 (z-score)	None	Mean (SD), g/m^3 Start of gender-affirming hormones: 0.26 (0.04) Age 22 years: 0.28 (0.05) No statistically significant difference z-score (SD) Start of gender-affirming hormones: -1.57 (1.74) Age 22 years: Not reported	Important	VERY LOW
Change from baseline in femoral neck BMAD in transfemales with a bone age less than 15 years ('young'; 24 months follow-up)									
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=16	None	Median (range), g/m^3 C0: 0.27 (0.20 to 0.33) C24: 0.27 (0.20 to 0.36)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							No statistically significant change z-score (range) C0: -1.32 (-3.39 to 0.21) C24: -1.30 (-3.51 to 0.92) No statistically significant change		
Change from baseline in femoral neck BMAD in transfemales with a bone age of 15 years or more ('old'; 24 months follow-up)									
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=6	None	Median (range), g/m^3 C0: 0.30 (0.26 to 0.34) C24: 0.29 (0.24 to 0.38) No statistically significant change z-score (range) C0: -0.36 (-1.50 to 0.46) C24: -0.56 (-2.17 to 1.29) No statistically significant change	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in femoral neck BMAD in transmales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=19 (Mean) N=18 (z-score)	None	Mean (SD), g/m^3 Start of gender-affirming hormones: 0.31 (0.04) Age 22 years: 0.33 (0.05) P=0.010 z-score (SD) Start of gender-affirming hormones: -0.28 (0.74) Age 22 years: Not reported	Important	VERY LOW
Change from baseline in femoral neck BMAD in transmales with a bone age of less than 14 years ('young'; 24 months follow-up)									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=10	None	Median (range), g/m^3 C0: 0.30 (0.22 to 0.35) C24: 0.33 (0.23 to 0.37) Statistically significant increase (p≤0.01) z-score (range) C0: -0.37 (-2.28 to 0.47) C24: -0.37 (-2.03 to 0.85) Statistically significant increase (p≤0.01)	Important	VERY LOW
Change from baseline in femoral neck BMAD in transmales with a bone age of 14 years or more ('old'; 24 months follow-up)									
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=23	None	Median (range), g/m^3 C0: 0.30 (0.23 to 0.41) C24: 0.32 (0.23 to 0.41) Statistically significant increase (p≤0.01) z-score (range) C0: -0.27 ((-1.91 to 1.29) C24: 0.02 (-2.1 to 1.35) Statistically significant increase (p≤0.05)	Important	VERY LOW
Change in lumbar spine BMD (2 uncontrolled, retrospective observational studies)									
Change from start of gender-affirming hormones to age 22 years in lumbar spine BMD in transfemales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=15 (Mean) N=13 (z-score)	None	Mean (SD), g/m^2 Start of gender-affirming hormones: 0.84 (0.11) Age 22 years: 0.93 (0.10) P<0.001 z-score (SD)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							Start of gender-affirming hormones: -1.01 (0.98) Age 22 years: -1.36 (0.83) No statistically significant difference		
Change from start of gender-affirming hormones to age 22 years in lumbar spine BMD in transmales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=19 (Mean and z-score)	None	Mean (SD), g/m^2 Start of gender-affirming hormones: 0.91 (0.10) Age 22 years: 0.99 (0.13) P<0.001 z-score (SD) Start of gender-affirming hormones: -0.72 (0.99) Age 22 years: -0.33 (1.12) No statistically significant difference	Important	VERY LOW
Change from start of testosterone treatment in lumbar spine BMD in transmen (follow-up 6 to 24 months)									
1 cohort study Stoffers et al. 2019	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable	N=62 (T0 and T6) N=37 (T12) N=15 (T24)	None	Mean (SD), g/cm^2 T0: 0.90 (0.11) T6: 0.94 (0.10) T12: 0.95 (0.09) T24: 0.95 (0.11) No statistically significant difference from T0 to any timepoint z-score (SD) T0: -0.81 (1.02) T6: -0.67 (0.95) T12: -0.66 (0.81) T24: -0.74 (1.17)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							No statistically significant difference from T0 to any timepoint		
Change in femoral neck BMD (2 uncontrolled, retrospective observational studies)									
Change from start of gender-affirming hormones to age 22 years in femoral neck BMD in transfemales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=15 (Mean) N=11 (z-score)	None	Mean (SD), g/m^2 Start of gender-affirming hormones: 0.87 (0.08) Age 22 years: 0.94 (0.11) P=0.009 z-score (SD) Start of gender-affirming hormones: -0.95 (0.63) Age 22 years: -0.69 (0.74) No statistically significant difference	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in femoral neck BMD in transmales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=19 (Mean) N=16 (z-score)	None	Mean (SD), g/m^2 Start of gender-affirming hormones: 0.88 (0.09) Age 22 years: 0.95 (0.10) P<0.001 z-score (SD) Start of gender-affirming hormones: -0.35 (0.79) Age 22 years: -0.35 (0.74) P=0.006	Important	VERY LOW
Change from start of testosterone treatment in right femoral neck (hip) BMD in transmales (follow-up 6 to 24 months)									
1 cohort study	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable	N=62 (T0 and T6)	None	Mean (SD), g/cm^2 T0: 0.77 (0.08)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Stoffers et al. 2019					N=37 (T12) N=15 (T24)		T6: 0.84 (0.11) T12: 0.82 (0.08) T24: 0.85 (0.11) No statistically significant difference from T0 to any timepoint z-score (SD) T0: -0.97 (0.79) T6: -0.54 (0.96) T12: -0.80 (0.69) T24: -0.31 (0.84) No statistically significant difference from T0 to any timepoint		
Change from start of testosterone treatment in left femoral neck (hip) BMD in transmales (follow-up 6 to 24 months)									
1 cohort study Stoffers et al. 2019	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable	N=62 (T0 and T6) N=37 (T12) N=15 (T24)	None	Mean (SD), g/cm^2 T0: 0.76 (0.09) T6: 0.83 (0.12) T12: 0.81 (0.08) T24: 0.86 (0.09) No statistically significant difference from T0 to any timepoint z-score (SD) T0: -1.07 (0.85) T6: -0.62 (1.12) T12: -0.93 (0.63) T24: -0.20 (0.70) No statistically significant difference from T0 to any timepoint	Important	VERY LOW

Abbreviations: BMAD: bone mineral apparent density; BMD: bone mineral density; g: grams; m: metre; SD: standard deviation

1 Downgraded 1 level - the cohort study by Klink et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding, no control group and high number of participants lost to follow-up)
2 Outcomes reported after gender reassignment surgery and not after gender-affirming hormones alone. Unclear whether observed changes are due to hormones or surgery
3 Downgraded 1 level - the cohort study by Vlot et al. (2017) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control)
4 Downgraded 1 level - the cohort study by Stoffers et al. (2019) was assessed as at high risk of bias (poor quality overall, lack of blinding and no control group)

Table 8: Question 2: For children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Cardiovascular risk factors

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change in body mass index (1 uncontrolled, retrospective observational study)									
Change from start of gender-affirming hormones to age 22 years in BMI in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI) +1.9 (0.6 to 3.2) Statistically significant increase (p<0.005) Mean BMI at 22 years (95% CI): 23.2 (21.6 to 24.8)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in BMI in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI) +1.4 (0.8 to 2.0) Statistically significant increase (p<0.005) Mean BMI at 22 years (95% CI): 23.9 (23.0 to 24.7)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Obesity rates at age 22 years (1 uncontrolled, retrospective observational study)									
Obesity rates at age 22 years in transfemales who started gender-affirming hormones as adolescents (1 uncontrolled, retrospective observational study)									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	At 22 years, 9.9% of transfemales were obese, compared with 3.0% in reference cisgender population No statistically analysis reported	Important	VERY LOW
Obesity rates at age 22 years in transfemales who started gender-affirming hormones as adolescents (1 uncontrolled, retrospective observational study)									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	At 22 years, 6.6% of transmales were obese, compared with 2.2% in reference cisgender population No statistically analysis reported	Important	VERY LOW
Change in blood pressure (1 uncontrolled, retrospective observational study)									
Change from start of gender-affirming hormones to age 22 years in systolic blood pressure (SBP) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI) -3 (-8 to 2) No statistically significant difference Mean SBP at 22 years (95% CI): 117 (113 to 122)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in diastolic blood pressure (DBP) in transfemales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI) +6 (3 to 10) Statistically significant increase (p<0.001) Mean DBP at 22 years (95% CI): 75 (72 to 78)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in systolic blood pressure (SBP) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI): +5 (1 to 9) Statistically significant increase (p<0.05) Mean SBP at 22 years (95% CI): 126 (122 to 130)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in diastolic blood pressure (DBP) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI): +6 (4 to 9) Statistically significant increase (p<0.001) Mean DBP at 22 years (95% CI): 74 (72 to 77)	Important	VERY LOW
Change in glucose levels, insulin levels, insulin resistance and HbA1c (2 uncontrolled, retrospective observational studies)									
Change from start of gender-affirming hormones to age 22 years in glucose level (mmol/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI): +0.1 (-0.1 to 0.2)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							No statistically significant difference Mean glucose level at 22 years (95% CI): 5.0 (4.8 to 5.1)		
Change from start of gender-affirming hormones to age 22 years in insulin level (mU/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI) +2.7 (-1.7 to 7.1) No statistically significant difference Mean insulin level at 22 years (95% CI): 13.0 (8.4 to 17.6)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in insulin resistance (HOMA-IR) in transfemales. Higher scores indicate more insulin resistance.									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI) +0.7 (-0.2 to 1.5) No statistically significant difference Mean HOMA-IR at 22 years (95% CI): 2.9 (1.9 to 3.9)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in glucose level (mmol/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI) 0.0 (-0.2 to 0.2) No statistically significant difference	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							Mean glucose level at 22 years (95% CI): 4.8 (4.7 to 5.0)		
Change from start of gender-affirming hormones to age 22 years in insulin level (mU/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI) -2.1 (-3.9 to -0.3) Statistically significant decrease (p<0.05) Mean insulin level at 22 years (95% CI): 8.6 (6.9 to 10.2)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in insulin resistance (HOMA-IR) in transmales. Higher scores indicate more insulin resistance.									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI): -0.5 (-1.0 to -0.1) Statistically significant decrease (p<0.05) Mean HOMA-IR at 22 years (95% CI): 1.8 (1.4 to 2.2)	Important	VERY LOW
Change from start of testosterone in HbA1c in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N= Not reported	None	No statistically significant change from start of testosterone treatment Numerical results, follow-up duration and further details of statistical analysis not reported.	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change in lipid profile (1 uncontrolled, retrospective observational study)									
Change from start of gender-affirming hormones to age 22 years in total cholesterol (mmol/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI): +0.1 (-0.2 to 0.4) No statistically significant difference Mean total cholesterol at 22 years (95% CI): 4.1 (3.8 to 4.4)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in HDL cholesterol (mmol/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI): 0.0 (-0.1 to 0.2) No statistically significant difference Mean HDL cholesterol at 22 years (95% CI): 1.6 (1.4 to 1.7)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in LDL cholesterol (mmol/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI): 0.0 (-0.3 to 0.2) No statistically significant difference Mean LDL cholesterol at 22 years (95% CI): 2.0 (1.8 to 2.3)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in triglycerides (mmol/L) in transfemales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI): +0.2 (0.0 to 0.5) Statistically significant increase (p<0.05) Mean triglycerides at 22 years (95% CI): 1.1 (0.9 to 1.4)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in total cholesterol (mmol/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI): +0.4 (0.2 to 0.6) Statistically significant increase (p<0.001) Mean total cholesterol at 22 years (95% CI): 4.6 (4.3 to 4.8)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in HDL cholesterol (mmol/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI) -0.3 (-0.4 to -0.2) Statistically significant decrease (p<0.001) Mean HDL cholesterol at 22 years (95% CI): 1.3 (1.2 to 1.3)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in LDL cholesterol (mmol/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI): +0.4 (0.2 to 0.6) Statistically significant increase (p<0.001)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							Mean LDL cholesterol at 22 years (95% CI): 2.6 (2.4 to 2.8)		
Change from start of gender-affirming hormones to age 22 years in triglycerides (mmol/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI) +0.5 (0.3 to 0.7) Statistically significant increase (p<0.001) Mean triglycerides at 22 years (95% CI): 1.3 (1.1 to 1.5)	Important	VERY LOW

Abbreviations: BMI: boss mass index; CI: confidence interval; DBP: diastolic blood pressure; HbA1c: glycated haemoglobin; HDL: high-density lipoproteins; HOMA-IR: Homeostatic Model Assessment of Insulin Resistance; LDL: low-density lipoproteins; mmol/L: millimoles per litre; mU/L: milliunits per litre; SBP: systolic blood pressure; SD: standard deviation

1 Downgraded 1 level - the cohort study by Klaver et al. (2020) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group)
2 Downgraded 1 level - the cohort study by Stoffers et al. (2019) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group)

Table 9: Question 2: For children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Other safety outcomes

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Liver enzymes (1 uncontrolled, retrospective observational study)									
Change from start of testosterone in aspartate aminotransferase (AST) level in transmales (up to 24 months follow-up)									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N= Not reported	None	No statistically significant change from start of testosterone treatment Numerical results, follow-up duration and further details of statistical analysis not reported.	Important	VERY LOW
Change from start of testosterone in alanine aminotransferase (ALT) level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N= Not reported	None	No statistically significant change from start of testosterone treatment Numerical results, follow-up duration and further details of statistical analysis not reported.	Important	VERY LOW
Change from start of testosterone in gamma-glutamyl transferase (GGT) level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N= Not reported	None	No statistically significant change from start of testosterone treatment Numerical results, follow-up duration and further details of statistical analysis not reported.	Important	VERY LOW
Change from start of testosterone in alkaline phosphatase (ALP) level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=62 (T0 and T1) N=37 (T12) N-15 (T24)	None	Median (IQR), U/L T0: 102 (78 to 136) T6: 115 (102 to 147) T12: 112 (88 to 143) T24: 81 (range 69 to 98) Statistically significant increase from T0 at T6 and T12 (p<0.001)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Kidney markers (1 uncontrolled, retrospective observational study)									
Change from start of testosterone in serum creatinine level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=62 (T0 and T1) N=37 (T12) N=15 (T24)	None	Mean (SD), umol/L T0: 62 (7) T6: 70 (9) T12: 74 (10) T24: 81 (10) Statistically significant increase from T0 at all timepoints (p<0.001)	Important	VERY LOW
Change from start of testosterone in serum urea[2] level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N= Not reported	None	No statistically significant change from start of testosterone treatment Numerical results, follow-up duration and further details of statistical analysis not reported.	Important	VERY LOW
Adverse effects (1 uncontrolled, retrospective observational study)									
Permanent discontinuation of gender-affirming hormones (median follow-up 2.0 years (range 0.0 to 11.3)									
1 cohort study Khatchadourian et al. 2014	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=63	None	No participants permanently discontinued gender-affirming hormones.	Important	VERY LOW
Temporary discontinuation of gender-affirming hormones (median follow-up 2.0 years (range 0.0 to 11.3)									
1 cohort study Khatchadourian et al. 2014	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=63	None	3/37 transmales receiving testosterone temporarily discontinued treatment, 2 due to concomitant mental health	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							comorbidities and 1 due to androgenic alopecia. All eventually resumed treatment. No transfemales receiving oestrogen temporarily discontinued treatment		
Minor complications during treatment with gender-affirming hormones (median follow-up 2.0 years (range 0.0 to 11.3)									
1 cohort study Khatchadourian et al. 2014	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=63	None	12/63 participants had minor complications during treatment with gender-affirming hormones All 12 were transmales receiving testosterone. Complications were severe acne (n=7), androgenic alopecia (n=1) mild dyslipidaemia (n=3) and significant mood swings (n=1) No transfemales receiving oestrogen had minor complications	Important	VERY LOW
Severe complications during treatment with gender-affirming hormones (median follow-up 2.0 years (range 0.0 to 11.3)									
1 cohort study Khatchadourian et al. 2014	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=63	None	No severe complications reported during gender-affirming treatment	Important	VERY LOW

Abbreviations: ALP: alkaline phosphatase; ALT: alanine aminotransferase; AST: aspartate aminotransferase; GGT: gamma-glutamyl transferase; IQR: interquartile range; SD: standard deviation; U/L: units per litre; umol/L: micromole per litre

1 Downgraded 1 level - the cohort study by Stoffers et al. (2019) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group)
2 Referred to as 'ureum' in original publication
3 Downgraded 1 level - the cohort study by Khatchadourian et al. (2014) was assessed as at high risk of bias (poor quality overall; lack of blinding, no control group and high number of participants lost to follow-up)

Table 10: From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria? – Transfemales compared with transmales

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Transfemales	Transmales	Result (95% CI)		
Impact on mental health (1 uncontrolled, retrospective observational study)									
Change from baseline in adjusted mean suicidality score, measured using the ASQ tool (mean treatment duration 349 days). Higher scores indicate a greater degree of suicidality.									
1 cohort study Allen et al. 2019	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=14	N=33	**Transfemales** T0 (baseline) = 1.21 (SE 0.36) T1 (final assessment) = 0.24 (SE 0.19) **Transmales** T0 (baseline) = 1.01 (SE 0.23) T1 (final assessment) = 0.29 (SE 0.13) No statistically significant difference in change from baseline between transfemales and transmales (p=0.79)	Critical	VERY LOW
Impact on quality of life (1 uncontrolled, retrospective observational study)									
Change from baseline in adjusted mean well-being score, measured using the GWBS of the Pediatric Quality of Life Inventory (mean treatment duration 349 days). Higher scores indicate better well-being.									
1 cohort study Allen et al. 2019	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=14	N=33	**Transfemales** T0 (baseline) = 58.44 (SE 4.09) T1 (final assessment) = 69.52 (SE 3.62) **Transmales** T0 (baseline) = 64.95 (SE 2.66)	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Transfemal es	Transmales	Result (95% CI)		
							T1 (final assessment) = 70.94 (SE 2.35) No statistically significant difference in change from baseline between transfemales and transmales (p=0.32)		

Abbreviations: ASQ: Ask Suicide-Screening Questions; GWBS: General Well-Being Scale; SE: standard error

1 The cohort study by Allen et al. 2019 was assessed at high risk of bias (poor quality; lack of blinding and no control group).

Table 11: From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria? – Sex assigned at birth males (transfemales)

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study type and number of studies Author year	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change from baseline in mean depression symptoms in transfemales, measured using the Quick Inventory of Depressive Symptoms (QIDS), self-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more depression.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=40	None	Baseline = 7.5 (SD 4.9) Follow-up = 6.6 (SD 4.4) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean depression symptoms in transfemales, measured using the Quick Inventory of Depressive Symptoms (QIDS), clinician-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe depression.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=45	None	Baseline = 4.2 (SD 3.2) Follow-up = 5.4 (SD 3.4) No statistical analysis reported for this sub-group	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study type and number of studies Author year	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change from baseline in mean anxiety symptoms in transfemales, measured using the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe anxiety.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=33	None	Baseline = 26.4 (SD 14.2) Follow-up = 24.3 (SD 15.4) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean panic symptoms in transfemales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=34	None	Baseline = 5.7 (SD 4.9) Follow-up = 5.1 (SD 4.9) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean generalised anxiety symptoms in transfemales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=34	None	Baseline = 8.6 (SD 5.1) Follow-up = 8.0 (SD 5.1) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean social anxiety symptoms in transfemales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=34	None	Baseline = 7.1 (SD 3.9) Follow-up = 6.8 (SD 4.4) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean separation anxiety symptoms in transfemales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=34	None	Baseline = 3.4 (SD 3.3) Follow-up = 2.7 (SD 2.3) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean school avoidance symptoms in transfemales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study type and number of studies Author year	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=33	None	Baseline = 1.8 (SD 1.7) Follow-up = 1.9 (SD 2.1) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in percentage of participants with suicidal ideation in transfemales, measured using the additional questions from the PHQ 9_Modified for Teens (approximately 12-month follow-up)									
1 cohort study Achille et al. 2020	Serious limitations[2]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=17	None	Wave 1 (baseline) = 11.8% (2/17) Wave 2 (approx. 12 months) = 5.9% (1/17) No statistical analysis reported	Critical	VERY LOW
Impact on body image (1 uncontrolled, prospective observational study)									
Change from baseline in mean body image in transfemales, measured using the BIS (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores represent a higher degree of body dissatisfaction.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=30	None	Baseline = 67.5 (SD 19.5) Follow-up = 49.0 (SD 21.6) No statistical analysis reported for this sub-group	Important	VERY LOW

Abbreviations: BIS: Body Image Scale; PHQ 9: Patient Health Questionnaire 9; SCARED: Screen for Child Anxiety Related Emotional Disorders; SD: standard deviation

1 Downgraded 1 level - the cohort study by Kuper et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
2 Downgraded 1 level - the cohort study by Achille et al. 2020 was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
3 Serious indirectness in Achille 2020- Approximately 30% of the full sample received puberty suppression alone or were receiving no treatment at final follow-up.

Table 12: From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria? – Sex assigned at birth females (transmales)

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change from baseline in mean depression symptoms in transmales, measured using the Quick Inventory of Depressive Symptoms (QIDS), self-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe depression.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=76	None	Baseline = 10.4 (SD 5.0) Follow-up = 7.5 (SD 4.5) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean depression symptoms in transmales, measured using the Quick Inventory of Depressive Symptoms (QIDS), clinician-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe depression.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=78	None	Baseline = 6.7 (SD 4.4) Follow-up = 6.2 (SD 4.1) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean anxiety symptoms in transmales, measured using the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe anxiety.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=65	None	Baseline = 35.4 (SD 16.5) Follow-up = 29.8 (SD 15.5) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean panic symptoms in transmales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=66	None	Baseline = 9.3 (SD 6.5) Follow-up = 7.9 (SD 6.5) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean generalised anxiety symptoms in transmales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=66	None	Baseline = 10.4 (SD 5.0) Follow-up = 9.0 (SD 5.1) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean social anxiety symptoms in transmales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=66	None	Baseline = 8.5 (SD 4.0) Follow-up = 7.8 (SD 4.1)	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Kuper et al. 2020							No statistical analysis reported for this sub-group		
Change from baseline in mean separation anxiety symptoms in transmales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=65	None	Baseline = 4.2 (SD 3.4) Follow-up = 3.4 (SD 2.6) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean school avoidance symptoms in transmales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=65	None	Baseline = 2.9 (SD 2.3) Follow-up = 2.0 (SD 2.3) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in percentage of participants with suicidal ideation in transmales, measured using the additional questions from the PHQ 9_Modified for Teens (approximately 12-month follow-up)									
1 cohort study Achille et al. 2020	Serious limitations[2]	Serious indirectness[3]	No serious inconsistency	Not calculable	N=33	None	Wave 1 (baseline) = 9.1% (3/33) Wave 2 (approx. 12 months) = 6.1% (2/33) No statistical analysis reported	Critical	VERY LOW
Impact on body image (1 uncontrolled, prospective observational study)									
Change from baseline in mean body image in transmales, measured using the BIS (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores represent a higher degree of body dissatisfaction.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=66	None	Baseline = 71.1 (SD 13.4) Follow-up = 52.9 (SD 16.8) No statistical analysis reported for this sub-group	Important	VERY LOW

Abbreviations: BIS: Body Image Scale; PHQ 9: Patient Health Questionnaire 9; SCARED: Screen for Child Anxiety Related Emotional Disorders; SD: standard deviation

1 Downgraded 1 level - the cohort study by Kuper et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).

2 Downgraded 1 level - the cohort study by Achille et al. 2020 was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
3 Serious indirectness in Achille 2020- Approximately 30% of the full sample received puberty suppression alone or were receiving no treatment at final follow-up.

Table 14: From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria? – Outcomes controlled for concurrent counselling and medicines for mental health problems

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Impact on mental health (1 uncontrolled, retrospective observational study)									
Change from baseline in mean depression score in transfemales, measured using the CESD-R (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicate more depression.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=17	None	No statistically significant change from baseline (p=0.27) Numerical scores not reported	Critical	VERY LOW
Change from baseline in mean depression score in transmales, measured using the CESD-R (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicate more severe depression.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=33	None	No statistically significant change from baseline (p=0.43) Numerical scores not reported	Critical	VERY LOW
Change from baseline in depression score in transfemales, measured using the Patient Health Questionnaire Modified for Teens (PHQ 9_Modified for Teens) (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicate more severe depression.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=17	None	No statistically significant change from baseline (p=0.07) Numerical scores not reported	Critical	VERY LOW
Change from baseline in depression score in transmales, measured using the Patient Health Questionnaire Modified for Teens (PHQ 9_Modified for Teens) (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicate more severe depression.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=33	None	No statistically significant change from baseline (p=0.67) Numerical scores not reported	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Impact on quality of life (1 uncontrolled, retrospective observational study)									
Change from baseline in mean quality of life score in transfemales, measured using the QLES-Q-SF (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicated better quality of life.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=17	None	No statistically significant change from baseline (p=0.06)	Critical	VERY LOW
Change from baseline in mean quality of life score in transmales, measured using the QLES-Q-SF (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicated better quality of life.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=33	None	No statistically significant change from baseline (p=0.08)	Critical	VERY LOW
Psychosocial Impact (1 uncontrolled, retrospective observational study)									
Functioning in adolescent development: Progresses normatively in school/ work during the real-life phase – impact on need for mental health treatment before or during gender identity assessment									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=49	None	Needed mental health treatment: 47% (15/32) functioning well Did not need mental health treatment: 82% (14/17) functioning well Statistically significant difference p=0.02	Important	VERY LOW
Functioning in adolescent development: Is age-appropriately able to deal with matters outside of the home during the real-life phase – impact on need for mental health treatment before or during gender identity assessment									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=49	None	Needed mental health treatment: 72% (23/32) managing well Did not need mental health treatment: 94% (16/17) managing well	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							No statistically significant difference p=0.06		
Functioning in adolescent development: Progresses normatively in school/ work during the real-life phase – impact on need for mental health treatment during the real-life phase									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=51	None	Needed mental health treatment: 42% (10/24) functioning well Did not need mental health treatment: 74% (20/27) functioning well Statistically significant difference p=0.02	Important	VERY LOW
Functioning in adolescent development: Is age-appropriately able to deal with matters outside of the home during the real-life phase – impact on need for mental health treatment during the real-life phase									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=51	None	Needed mental health treatment: 67% (16/24) managing well Did not need mental health treatment: 93% (25/27) managing well Statistically significant difference p=0.02	Important	VERY LOW

Abbreviations: CESD-R: Center for Epidemiologic Studies Depression; p: p-value; PHQ 9: Patient Health Questionnaire 9; QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire

1 Downgraded 1 level - the cohort study by Achille et al 2020 was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
2 Serious indirectness in Achille 2020- Approximately 30% of the full sample received puberty suppression alone or were receiving no treatment at final follow-up.
3 Downgraded 1 level - the cohort study by Kaltiala et al. 2020 was assessed at high risk of bias (poor quality; lack of blinding and no control).

Table 15: From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria? – Tanner age

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Impact on mental health (1 uncontrolled, retrospective observational study)									
Change from baseline in mental health problems – depression, anxiety and anxiety-related symptoms (mean duration of gender-affirming hormone treatment was 10.9 months)									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=105	None	No difference in outcomes found by Tanner age. Numerical results, statistical analysis and information on specific outcomes not reported. It is unclear from the paper whether Tanner age is at initial assessment, start of GnRH analogues, start of gender-affirming hormones, or another timepoint	Critical	VERY LOW
Impact on body image (1 uncontrolled, prospective observational study)									
Change from baseline in mean body image, measured using the BIS (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores represent a higher degree of body dissatisfaction.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=105	None	No difference in body image score found by Tanner age. Numerical results, statistical analysis and information on specific outcomes not reported. It is unclear from the paper whether Tanner age is at initial assessment, start of GnRH analogues, start of gender-affirming hormones, or another timepoint	Important	VERY LOW

Abbreviations: BIS: Body Image Scale

1 Downgraded 1 level - the cohort study by Kuper et al. 2020 was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).

Glossary

Ask Suicide-Screening Questions (ASQ)	ASQ is a four-item dichotomous (yes, no) response measure with high sensitivity, designed to identify risk of suicide. A patient is considered to have screened positive if they answered yes to any item. The authors of Allen et al. 2019 altered the fourth item of the ASQ ("Have you ever tried to kill yourself?") and prefaced it with "In the past few weeks . . ." as they were not investigating lifetime suicidality. A response of 'no' was scored as 0 and a response of 'yes' was scored as 1; each item was summed, generating an overall score for suicidality on a scale ranging from 0 to 4, with higher scores indicating greater levels of suicidal ideation.
Beck Depression Inventory-II (BDI-II)	The BDI-II is a tool for assessing depressive symptoms. There are no specific scores to categorise depression severity, but it is suggested that 0 to 13 is minimal symptoms, 14 to 19 is mild depression, 20 to 28 is moderate depression, and severe depression is 29 to 63.
Body Image Scale (BIS)	The BIS is used to measure body satisfaction. The scale consists of 30 body features, which the person rates on a 5-point scale. Each of the 30 items falls into one of 3 basic groups based on its relative importance as a gender-defining body feature: primary sex characteristics, secondary sex characteristics, and neutral body characteristics. A higher score indicates more dissatisfaction.
Bone mineral apparent density (BMAD)	BMAD is a size adjusted value of bone mineral density (BMD) incorporating bone size measurements using UK norms in growing adolescents.
Center for Epidemiologic Studies Depression scale (CESD-R)	The CESD-R is a valid, widely used tool to access depressive symptoms. The CESD-R asks about how frequently a person has felt or behaved in a certain way; with 20 questions scored from 0 score is calculated as a sum of 20 questions, ranging from 0 ("not at all or less than one day") to 3 ("5–7 days" and/or "nearly every day for 2 weeks"). Total score ranges from 0 to 60, with higher scores indicating more depressive symptoms.
Cisgender	Cisgender is a term for someone whose gender identity matches their birth-registered sex.
Family APGAR (Adaptability, Partnership, Growth, Affection and Resolve) test	The Family APGAR test is a 5-item questionnaire, with higher scores indicating better family functioning. The authors reported the following interpretation of the score: functional, 17-20 points; mildly dysfunctional, 16-13 points; moderately dysfunctional, 12-10 point; severely dysfunctional, <9 points.
Gender	The roles, behaviours, activities, attributes and opportunities that any society considers appropriate for girls and boys, and women and men.
Gender dysphoria	Discomfort or distress that is caused by a discrepancy between a person's gender identity (how they see themselves regarding their gender) and that person's sex assigned at birth (and the associated gender role, and/or primary and secondary sex characteristics).

General Well-Being Scale (GWBS) of the Pediatric Quality of Life Inventory score	The GWBS of the Pediatric Quality of Life Inventory uses uses a 5-point response scale, contains seven items, and measures two dimensions: general wellbeing (6 items) and general health (1 item). Each item is scored from 0 to 4, and the total score is linearly transformed to a 0 to 100 scale. High scores reflect fewer perceived problems and greater well-being.
GnRH analogue	GnRH analogues competitively block GnRH receptors to prevent the spontaneous release of two gonadotropin hormones, Follicular Stimulating Hormone (FSH) and Luteinising Hormone (LH) from the pituitary gland. The reduction in LH and FSH secretion reduces oestradiol secretion from the ovaries in those whose sex assigned at birth was female and testosterone secretion from the testes in those whose sex assigned at birth was male.
Patient Health Questionnaire Modified for Teens score (PHQ 9_Modified for Teens)	The PHQ 9_Modified for Teens is a validated tool to assess depression, dysthymia and suicide risk. The tool consists of 9 questions scored from 0 to 3 (total score 0 to 27), plus an additional 4 questions that are not scored. A score of 0 to 4 suggests no or minimal depressive symptoms, 5 to 9 mild, 10-14 moderate, 15-19 moderate and 20-27 severe symptoms.
Quick Inventory of Depressive Symptoms (QIDS)	Both the clinician- and self-reported QIDS are validated tools to assess depressive symptoms. The tool consists of 16 items, with the highest score for 9 items (sleep, weight, psychomotor changes, depressed mood, decreased interest, fatigue, guilt, concentration, and suicidal ideation) are added to give a total score ranging from 0 to 27. A score of 0 to 5 is suggestive of no depressive symptoms, 6 to 10 mild symptoms, 11 to 15 moderate symptoms, 16-20 severe symptoms and 21 to 27 very severe symptoms.
Quality of Life Enjoyment and Satisfaction Questionnaire (QLES-Q-SF)	QLES-Q-SF is a validated questionnaire, consisting of 15 questions that rate quality of life on a scale of 1 (poor) to 5 (very good).
Screen for Child Anxiety Related Emotional Disorders (SCARED) questionnaire	SCARED is a validated, 41-point questionnaire, with each item scored 0 to 2. A total score of 25 or more is suggestive of anxiety disorder, with scores above 30 being more specific. Certain scores for specific questions may indicate the presence of other anxiety-related disorders: A score of 7 or more in questions related to panic disorder or significant somatic symptoms may indicate the presence of these. A score of 9 or more in questions related to generalised anxiety disorder may indicate the presence of this. A score of 5 or more in questions related to separation anxiety may indicate the presence of this. A score of 8 or more in questions related to social anxiety disorder may indicate the presence of this. A score of 3 or more in questions related to significant school avoidance may indicate the presence of this.
State-Trait Anxiety Inventory (STAI) score	STAI is a validated and commonly used measure of state anxiety (current state of anxiety) and trait anxiety (general state of calmness, confidence and security). It has 40 items, the first 20 covering state anxiety, the second 20 covering trait anxiety. STAI

	can be used in clinical settings to diagnose anxiety and to distinguish it from depressive illness. Each subtest (state and trait) is scored between 20 and 80, with higher scores indicating greater anxiety. There is no published minimal clinically meaningful difference (MCID) for STAI or thresholds for anxiety severity.
Strengths and Difficulties Questionnaire (SDQ, Spanish version	The SDQ, Spanish version includes 25-items covering emotional symptoms, conduct problems, hyperactivity/ inattention, peer relationship problems and prosocial behaviour. The authors state that a score of more than 20 is considered indicative of risk of having a disorder (normal: 0-15; borderline: 16-19, abnormal: 20-40).
Tanner stage	Tanner staging is a scale of physical development.
Transgender (including transmale and transfemale)	Transgender is a term for someone whose gender identity is not congruent with their birth-registered sex. A transfemale is a person who identifies as female and a transmale is a person who identifies as male.
Utrecht Gender Dysphoria Scale (UGDS)	The UGDS is a validated screening tool for both adolescents and adults to assess gender dysphoria. It consists of 12 items, to be answered on a 1- to 5-point scale, resulting in a sum score between 12 and 60. Higher scores indicate higher levels of gender dysphoria.

References

Included studies

- Achille, C., Taggart, T., Eaton, N.R. et al. (2020) Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: Preliminary results. International Journal of Pediatric Endocrinology 2020(1): 8
- Allen, LR, Watson, LB, Egan, AM et al. (2019) Well-being and suicidality among transgender youth after gender-affirming hormones. Clinical Practice in Pediatric Psychology 7(3): 302-311
- Kaltiala, R., Heino, E., Tyolajarvi, M. et al. (2020) Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria. Nordic Journal of Psychiatry 74(3): 213-219
- Khatchadourian K, Amed S, Metzger DL (2014) Clinical management of youth with gender dysphoria in Vancouver. The Journal of pediatrics 164(4): 906-11
- Klaver, Maartje, de Mutsert, Renee, van der Loos, Maria A T C et al. (2020) Hormonal Treatment and Cardiovascular Risk Profile in Transgender Adolescents. Pediatrics 145(3)
- Klink D, Caris M, Heijboer A et al. (2015) Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. The Journal of Clinical Endocrinology and Metabolism 100(2): e270-5

- Kuper, Laura E, Stewart, Sunita, Preston, Stephanie et al. (2020) Body Dissatisfaction and Mental Health Outcomes of Youth on Gender-Affirming Hormone Therapy. Pediatrics 145(4)
- Lopez de Lara, D., Perez Rodriguez, O., Cuellar Flores, I. et al. (2020) Psychosocial assessment in transgender adolescents. Anales de Pediatria
- Stoffers, Iris E; de Vries, Martine C; Hannema, Sabine E (2019) Physical changes, laboratory parameters, and bone mineral density during testosterone treatment in adolescents with gender dysphoria. The journal of sexual medicine 16(9): 1459-1468
- Vlot MC, Klink DT, den Heijer M et al. (2017) Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone 95: 11-19

Other references

- World Health Organisation (2018) International Classification of Diseases 11. Available from https://icd.who.int/ [accessed 27 August 2020]
- American Psychiatric Association. (2013). Diagnostic and statistical Manual of Mental Disorders (DSM-5) (5th ed). Washington, DC and London: American Psychiatric Publishing. pp.451-460. Available from: https://www.psychiatry.org/patients-families/gender-dysphoria/what-is-gender-dysphoria [accessed 27 August 2020]
- NHS England (2013). NHS Standard contract for gender identity development service for children and adolescents https://www.england.nhs.uk/wp-content/uploads/2017/04/gender-development-service-children-adolescents.pdf [accessed 27 August 2020]
- NHS England (2016). Clinical Commissioning Policy: Prescribing of Cross-Sex Hormones as part of the Gender Identity Development Service for Children and Adolescents https://www.england.nhs.uk/wp-content/uploads/2018/07/Prescribing-of-cross-sex-hormones-as-part-of-the-gender-identity-development-service-for-children-and-adolesce.pdf [accessed 27 August 2020]

Copyright

Received: 20 January 2023 | Revised: 5 April 2023 | Accepted: 17 April 2023

DOI: 10.1111/apa.16791

ACTA PÆDIATRICA WILEY

REVIEW ARTICLE

A systematic review of hormone treatment for children with gender dysphoria and recommendations for research

Jonas F. Ludvigsson[1,2,3] | Jan Adolfsson[4,5] | Malin Höistad[5] | Per-Anders Rydelius[6,†] | Berit Kriström[7] | Mikael Landén[1,8]

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[2]Department of Paediatrics, Örebro University Hospital, Örebro, Sweden
[3]Division of Digestive and Liver Disease, Department of Medicine, Columbia University Medical Center, New York, New York, USA
[4]Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden
[5]The Swedish Agency for Health Technology Assessment and Assessment of Social Services, Stockholm, Sweden
[6]Department of Women's and Children's health, Karolinska Institutet, Stockholm, Sweden
[7]Department of Clinical Sciences/Paediatrics, Umeå University, Umeå, Sweden
[8]Department of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden

Correspondence
Mikael Landén, Section of Psychiatry and Neurochemistry, Sahlgrenska University Hospital, Blå Stråket 15, S-431 45 Gothenburg, Sweden.
Email: mikael.landen@neuro.gu.se

Berit Kriström, Department of Clinical Sciences/Paediatrics, Umeå University, S-90185 Umeå, Sweden.
Email: berit.kristrom@umu.se

Funding Information
Swedish Agency for Health Technology Assessment and Assessment of Social Services

Abstract

Aim: The aim of this systematic review was to assess the effects on psychosocial and mental health, cognition, body composition, and metabolic markers of hormone treatment in children with gender dysphoria.

Methods: Systematic review essentially follows PRISMA. We searched PubMed, EMBASE and thirteen other databases until 9 November 2021 for English-language studies of hormone therapy in children with gender dysphoria. Of 9934 potential studies identified with abstracts reviewed, 195 were assessed in full text, and 24 were relevant.

Results: In 21 studies, adolescents were given gonadotropin-releasing hormone analogues (GnRHa) treatment. In three studies, cross-sex hormone treatment (CSHT) was given without previous GnRHa treatment. No randomised controlled trials were identified. The few longitudinal observational studies were hampered by small numbers and high attrition rates. Hence, the long-term effects of hormone therapy on psychosocial health could not be evaluated. Concerning bone health, GnRHa treatment delays bone maturation and bone mineral density gain, which, however, was found to partially recover during CSHT when studied at age 22 years.

Abbreviations: BMD, bone mineral density; CSHT, cross-sex hormone treatment; DXA, dual-energy X-ray absorptiometry; GnRHa, gonadotropin-releasing hormone agonist (analogues); GRADE, grades of recommendation, assessment, development and evaluation; ICD, International Classification of Diseases; MRI, magnetic resonance imaging; SBU, Swedish Agency for Health Technology Assessment and Assessment of Social Services.

Berit Kriström and Mikael Landén have equal contribution.

†Part of the original study group but deceased in December 2021.

Conclusion: Evidence to assess the effects of hormone treatment on the above fields in children with gender dysphoria is insufficient. To improve future research, we present the GENDHOR checklist, a checklist for studies in gender dysphoria.

KEYWORDS
adolescent, bone density, gender dysphoria, gonadotropin-releasing hormone agonist, psychosocial functioning

1 | INTRODUCTION

Gender incongruence refers to a mismatch between the biological sex and perceived gender identity. When gender incongruence causes significant discomfort, it is called gender dysphoria. When gender dysphoria causes clinically significant distress, the condition might meet the diagnostic criteria for transsexualism according to the (international classification of disease) ICD-10 guidelines,[1] or gender dysphoria according to the DSM-5.[2] Gender identity-affirming health care is provided to ease gender dysphoria.[3] The treatment aims to align bodily characteristics with the individual's gender identity, and usually includes cross-sex hormone treatment (CSHT), as well as chest and genital surgery.

In youth with gender dysphoria, gonadotropin-releasing hormone analogues (GnRHa) have been used to inhibit spontaneous puberty development. The rationale is to prevent irreversible bodily changes and give young individuals time to explore their gender identity. Following the first case report in which a GnRHa was used to suppress puberty in a female-to-male transsexual individual,[4] the "Dutch protocol" was developed.[5] According to this protocol, young pubertal people presenting with gender dysphoria should first undergo a thorough psychological evaluation. If the diagnosis gender dysphoria is confirmed, GnRHa treatment is recommended to start during the early stages of puberty (Tanner stages 2–3). If gender dysphoria subsides, the individual may discontinue GnRHa treatment, at which point spontaneous puberty will restart. If gender dysphoria persists, CSHT might start at age 16 years and sex-reassignment surgery at 18 years. Gender dysphoria in youth was a rare phenomenon when the Dutch multidisciplinary protocol for the treatment of gender dysphoria was introduced. Seeking care for gender dysphoria has since become increasingly common in younger people in many parts of the western world,[6,7] with an exponential rise among children born female.[8] Although not all children with gender dysphoria receive gender identity affirming treatment, there has been an ensuing increase in hormones to treat children with gender dysphoria, of which data on the effects and side effects are limited. There is no previous systematic review or meta-analysis of hormone treatment for children with gender dysphoria.

This systematic review aimed at assessing (a) psychosocial effects, (b) effects on bone health, (c) effects on body composition and metabolism, and (d) satisfaction and therapy persistence in children aged <18 years with gender dysphoria undergoing hormone therapy.

Key Notes

- This systematic review assessed psychosocial effects, bone health, body composition and metabolism, and therapy persistence in children (<18 years of age) with gender dysphoria undergoing treatment with gonadotropin-releasing hormone analogues (GnRHa).
- Long-term effects of hormone therapy on psychosocial health are unknown. GnRHa treatment delays bone maturation and gain in bone mineral density.
- GnRHa treatment in children with gender dysphoria should be considered experimental treatment of individual cases rather than standard procedure.

In this review, trans women are referred to as male-to-female and trans men as female-to-male.

2 | METHODS

2.1 | Preregistration

This systematic review originated from a 2-year commissioned work from the governmental body the Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU). Ongoing SBU reviews are registered on the SBU website (https://www.sbu.se/en/ongoing-projects/) but not recorded in external databases.

2.2 | Selection criteria

The search was restricted to children aged <18 years with reported gender dysphoria. We included observational studies, randomised controlled trials, and systematic reviews according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[9] Case reports, editorials, and non-human studies were excluded from further review. The search was limited to English-language publications.

2.3 | Search strategy

Two professional information specialists at the Swedish Agency for Health Technology Assessment and Assessment for Social Services (SBU) performed a comprehensive search of the following medical databases up until 9 November 2021: CINAHL (EBSCO), Cochrane Library (Wiley), EMBASE (Embase.com), PsycINFO (EBSCO), PubMed (NLM), Scopus (Elsevier), and SocINDEX (EBSCO). They also searched the Campbell Library, Epistemonikos, Evidence Search, International HTA database, as well as three NIHR Centre for Reviews and Dissemination (CRD) databases: Database of Abstracts of Reviews of Effects (DARE), Health, and Technology Assessment (HTA), and NHS Economic Evaluation Database (EED). Finally, we searched PROSPERO, an international prospective register for systematic reviews, to identify any relevant ongoing systematic reviews but found none. The search, selection, and assessment were conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.[9] The search and selection processes are outlined in Figure 1. Only studies of low or moderate bias were eligible for this review. Full literature search strategy is provided at the SBU web page (https://www.sbu.se/contentassets/4062b596a35c4e1383405766b7365076/bilaga-1-litteratursokning.pdf).

2.4 | Relevance, risk of bias, and quality of evidence

Two independent experts checked all hits for relevance. Relevant studies (based on a pre-defined PICO) were then evaluated for risk of bias, also by two independent experts, according to ROBINS-I (Risk of bias in non-randomised studies of interventions).[10,11] Robins-I assesses possible bias in seven domains: confounding, bias due to selection, measurement classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, and selection of the reported result.

If the two reviewers did not agree on content or quality, the paper was discussed in the larger research team of four experts (JFL, PR, BK, ML). Randomised controlled trials were planned to be assessed by RoB-2.[10,11] To rate the quality of evidence for specific outcomes, we used the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) system.[12] GRADE has four levels of evidence (very low, low, moderate, high) and considers five domains that can decrease the level of certainty one or two levels (risk of bias, imprecision, inconsistency, indirectness (similar to 'external validity'), and publication bias).

2.5 | Data extraction

Two reviewers (MH, JA) retrieved data from the included studies. The data extracted included the outcomes mental and psychosocial health including suicidality, anthropometric measures and metabolism, bone health, adverse events, and the characteristics of each study including age at referral or intake, age at start of GnRHa treatment, age at start of CSHT, number of participants enrolled in study, number of transgender participants, number of hormone treated transgender participants, number of non-transgender participants, number of participants evaluated, treatment type (drugs, dosages, type of administration, treatment frequency), total treatment duration, and total follow-up time. The full data extraction of included studies is provided at the SBU web page (https://www.sbu.se/contentassets/4062b596a35c4e1383405766b7365076/bilaga-3-tabellverk-over-inkluderade-studier.pdf).

2.6 | Statistics

No statistical analyses were performed.

2.7 | Ethics

Ethical approval is not applicable for this systematic review.

3 | RESULTS

3.1 | Identified studies

After duplicate removal, the search yielded 9934 potential studies (Figure 1). Of these, 195 were selected for thorough reading. Of these, 36 were relevant and assessed for risk of bias. Twelve studies were excluded because of high risk for bias, leaving 24 studies with low to moderate, moderate, or moderate to high risk of bias reviewed in this paper. A list of excluded studies is provided at the SBU web page (https://www.sbu.se/contentassets/4062b596a35c4e1383405766b7365076/bilaga-2-exkluderade-studier-med-hog-risk-for-bias.pdf).

3.2 | Characteristics of the 24 studies

All 24 relevant studies had been published since 2014 (Table 1). Study participant age at the start of GnRHa therapy was typically between 11 and 15 years (range 9–18.6 years), with CSHT rarely being introduced before age 15. Except for the Hisle-Gorman et al.[6] ($n = 3754$ participants) and Mullins et al.[13] ($n = 611$) papers, few studies included >200 individuals. GnRHa treatment often continued for around 2 years, sometimes up to 4 years, and similar treatment durations were observed or reported for CSHT as observations were usually not reported after age 18 years. Full details of included studies are given at the SBU web page. Overall, there were eight studies on GnRH alone, 13 studies on GnRH + CSHT, and three studies on CSHT alone.

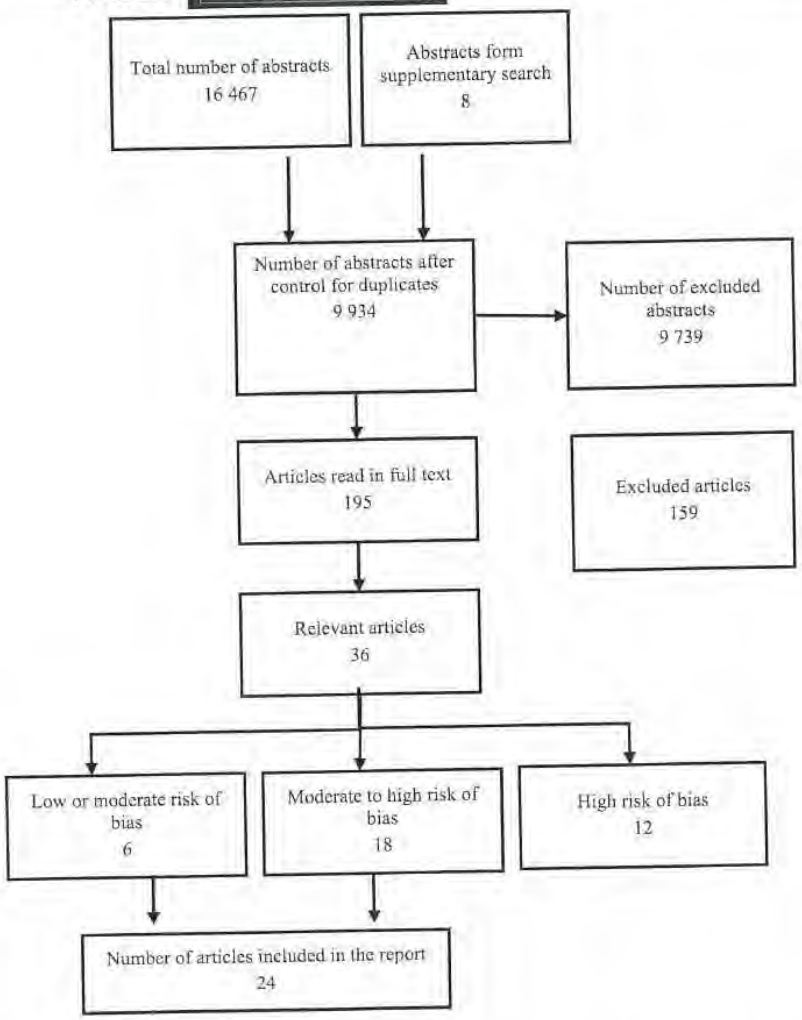


FIGURE 1 PRISMA flow diagram.

3.3 | Psychosocial and mental health

Table 2 outlines the six studies that examined psychosocial outcomes and cognitive effects.[14–19] Three of these studies found significantly improved overall psychosocial function after GnRHa treatment as measured by the Children's Global Assessment Scale (CGAS).[14–16] Two of these studies observed no statistically significant change in gender dysphoria.[15,16] Two of these studies reported significantly improved self-rated quality of life after treatment measured through Kidscreen-27, Short Form-8 (SF-8), Child Behaviour Checklist (CBCL) (parent report), and Youth Self Report (YSR),[16,17] while another study reported no statistically significant differences in anxiety and depression between those who started and not started hormone therapy.[18]

Because these studies were hampered by small number of participants and substantial risk of selection bias, the long-term effects of hormone treatment on psychosocial health could not be evaluated. Of note, the above studies do not allow separation of potential effects of psychological intervention independent of hormonal effects.

3.4 | Cognitive outcomes

We could only identify one study of low-moderate bias on cognitive outcomes in children with gender dysphoria receiving GnRHa therapy.[19] This cross-sectional study from the USA comprised 20 treated (8 male-to-female and 12 female-to-male) and 20 untreated (10 male-to-female and 10 female-to-male) young transgender persons and a control group ($n=45$). Controls were identified from age-matched family members and friends. The Tower of London task was administered to assess executive functioning. The study neither found differences in cognitive function between treated and untreated transgender persons, nor between treated transgender persons and controls. However, because no before-after GnRHa therapy analyses were performed, the study

TABLE 1 Overview of 24 included studies.

Reference	Ages of patients (years)			Numbers of patients						Interventions			Time: duration and follow-up			Outcomes extracted
	Age at intake range (mean)	Age at start of GnRH range (mean)	Age at start of CSHT range (mean)	*n* referred	*n* TG enrolled	*n* TG HT	*n* TG non-HT	*n* non-TG	*n* TG HT at last FU	GnRH	CSHT	Surgery[b]	GnRH duration range (mean)	CSHT duration range (mean)	Follow-Up time range (mean)	Mental health Bone health Anthropometrics Metabolism
Mental health																
de Vries 2014[14]	11–17 (13.6)	11.5–18.5 (14.8)	13.9–19 (16.7)	196	111	55			32	x	x	x	1 year[a]	4 years[a]		UGDS, global functioning (CGAS), depression (BDI), anxiety (STAI), anger (TPI)
Costa 2015[15]	12–17 (15.5)	13–17 (16.5)		436	201	101	100		35	x			1 year		1.5 years	UGDS, psychosocial functioning (CGAS)
Becker-Hebly 2020[17]		11–17 (15.5)	13–17 (15.5)	434	75	54	21		54	x	x	x	0.5–4 years[a]	0.5–4 years[a]	7–49 months	Global functioning (CGAS), psychosocial functioning (YSR/ASR)
Cantu 2020[18]		11–xx (15)	xx–18 (15)		80	42	38		28	x	x		NR	NR	1–11 months (5 months)	Psychosocial functioning (PHQ-9, GAD-7), acute distress, suicidality
Carmichael 2021[16]		12.0–15.3 (13.6)			44	44			14	x			12–59 months (31 months)		12–36 months	UGDS, CGAS, psychological functioning (CBCL, YSR), Self-harm, BIS, HRQoL (Kidscreen52)
Hisle-Gorman 2021[6]	8–13 (10)		16.6–19.8 (18.2)		3754	963		6603	963	x	x		0.7–2.7 years (1.5)	0.7–2.7 years (1.5)	8.5 years	Mental health diagnosis, psychotropic medication use, medication days, service use
Staphorsius 2015[19]		min 12			41	20	20	45		x			0.6–2.6 years (1.6)			Psychological functioning (CBCL), cognitive function (executive function task)

(Continues)

TABLE 1 (Continued)

Reference	Ages of patients (years)			Numbers of patients						Interventions			Time: duration and follow-up			Outcomes extracted
	Age at intake range (mean)	Age at start of GnRH range (mean)	Age at start of CSHT range (mean)	*n* referred	*n* TG enrolled	*n* TG HT	*n* TG non-HT	*n* non-TG	*n* TG HT at last FU	GnRH	CSHT	Surgery[b]	GnRH duration range (mean)	CSHT duration range (mean)	Follow-Up time range (mean)	Mental health Bone health Anthropometrics Metabolism
Bone health																
Joseph 2019[23]		12–14 (13)				70			70	x			1–xx years		up to 2.8 years	Height, weight, BMI BMD, BMAD, Z-score (hip, spine)
Klink 2015[21]		11.4–18.3 (15)	15.6–19 (16)			34			34	x	x	x	0.25–8 years	xx–8 years	up to age 22	Height, BMD, aBMD, Z-score, T-score (femoral neck, lumbar spine)
Vlot 2017[22]		11.5–18.6 (14)	14.0–19.5 (16)		215	70			57	x	x		1–xx years		up to 2 years	Height, BMAD, Z-score (hip, lumbar spine), bone markers (P1NP, OC, ICTP)
Schagen 2020[20]		12.2–16.5 (14)	15.0–17.9 (16)			127			121	x	x		1.5–4 years	3 years		aBMD, Z-score (hip)
Stoffers 2019[24]		11.8–18.0 (16)	14.9–18.4 (17.2)		64	62			15	x	x		3 months–3 years	5 months–3 years	2 years	Height, BP, BMD, Z-score (femoral neck, lumbar spine)
Navabi 2021[25]		13.4–17.4 (15)			198	172			116	x			6 months–2 years		1.5 years	BMD, aBMAD, Z-score (hip, lumbar spine)
van der Loos 2021[26]		11–17	15–17			322			322	x	x	x	1–3 years	2–6 years	up to 4 years	Subperiostal width, endocortical diameter
Lee 2020[27]		9.6–13.4 (11.5)			95	63			63	x			2 months			BMD, aBMAD, Z-score (hip, lumbar spine)
Anthropometrics and metabolism																
Schagen 2016[28]		11.1–18.6 (14)			138	116			77	x			3–12 months		1 year	Height, weight, BMI, lean body mass, liver enzymes, creatinine
Klaver 2018[31]		12.7–17.3[a] (15)	15.3–17.8[a] (16)	489	192	192			192	x	x	x	0.5–2.9 years (1.5[a])	1.6–3.4 years (2.9[a])	age 22	Weight, BMI, total body %, WHR

TABLE 1 (Continued)

	Ages of patients (years)			Numbers of patients						Interventions			Time: duration and follow-up			Outcomes extracted
Reference	Age at intake range (mean)	Age at start of GnRH range (mean)	Age at start of CSHT range (mean)	*n* referred	*n* TG enrolled	*n* TG HT	*n* TG non-HT	*n* non-TG	*n* TG HT at last FU	GnRH	CSHT	Surgery[b]	GnRH duration range (mean)	CSHT duration range (mean)	Follow-Up time range (mean)	Mental health Bone health Anthropometrics Metabolism
Klaver 2020[32]		12.8–17.2[a] (14.9)	15.3–17.8[a] (16.6)		192	192			192	x	x	x	0.5–2.9 years (1.5)[a]	1.1–3.4 years (2.5[a])	age 22	BMI, SBP, DBP, glucose, insulin, HOMA-IR, cholesterol, triglycerides
Peri 2020[33]		13.4–15.4 (14)	14.2–16.0 (15)		48	15			15	x	x		2–4 months	2–6 months		BMI, BP
Schulmeister 2021[29]		9.0–14.5 (11.5)			92	55	226		55	x			10–14 months		1 year	Height velocity, BMI, z-score
Nokoff 2021[30]		10.2–14.1 (12)			17	17	31		17	x			0.5–5.8 years			Insulin, glucose HbA1c, HOMA-IR, body fat, % lean mass
Tack 2016[34]			NR (15–17)		45	43			43		x			6–18 months (12)	1.5 years	Height, weight, BMI, triglycerides, cholesterol, suicide, side effects
Jarin 2017[35]		10.3–xx	xx–25 (16–18)		116	116			116	(x)	x				2 years	BMI, BP, haematocrit, Hb, cholesterol
Mullins 2021[13]			13–24 (17)	1406	611	611			611		x			0.8–2.8 years (1.5 years)	3 years	Haematology, thrombosis, BMI

Note: Number of patients:*n* referred = number of patients referred to gender clinic for evaluation of gender dysphoria (not same at number of patients receiving GD diagnosis)*n* TG enrolled = number of patients enrolled in the study at start*n* TG = number of patients with gender dysphoria TG HT = number of patients with gender dysphoria treated with hormones (GnRH alone, GnRH + CSHT, or CSHT only)*n* TG non-HT = number of patients with gender dysphoria treated NOT with hormones*n* TG HT at last FU = number of patients with gender dysphoria treated with hormones (GnRH alone, GnRH + CSHT, or CSHT only) evaluated at last follow-up time*n* non-TG = number of subjects in study without gender dysphoria (reference population).

Abbreviations: BDI, Beck Depression Inventory; BIS, Body Image Scale; BMAD, Bone Mineral Apparent Density; BMD, Bone Mineral Density; BMI, Body Mass Index; BP, Blood pressure; CBCL, Child Behaviour Checklist; CGAS, Global functioning Children's Global Assessment Scale, [higher scores (>80) indicating better global functioning]; CSHT, Cross-Sex Hormone Treatment/ gender-affirming treatment, testosterone, oestradiol, cyproterone acetate (CA), spironolactone, lynestrenol; GAD-7, Generalised Anxiety Disorder-7; GnRH, Gonadotropin Releasing Hormone analogue: triptorelin; HRQoL, Health Related Quality of Life; HT Hormone treatment, either GnRH, CSHT, or both; PHQ-9, Patient Health Questionnaire-9; SF-8, Short Form-8: (<18 years); STAI, Spielberger's Trait Anxiety; TG, Transgender; TPI, Anger Spielberger's Trait Anger; UGDS, Utrecht Gender Dysphoria Scale, score range 12–60 points [high score = high level of GD]; WHR, Waist-hip ratio; YSR, Youth Self Report; YSR (ages 11–18 years); Adult version (ASR, >18 years), [higher scores reflect higher degree of problems]; NR, not reported.

[a]Calculated by SBU.

[b]Surgery = any kind of gender reassignment surgery (gonadectomy, mastectomy, hysterectomy, laryngeal surgery, hair removal, phalloplasty, vaginoplasty).

TABLE 2 Summary of findings on psychosocial outcomes of puberty-blocking treatment (GnRHa) treatment in children with gender dysphoria.[14–17]

Outcome measures	Number of study participants, description of studies	Main result	"Certainty of evidence"	Deduction in GRADE[a]
Global function	n on hormones = 254 n evaluated = 113 Four observational cohort studies: one prospective and three retrospective studies[14–17]	Improved global function as assessed with the CGAS	Cannot be assessed	−2 risk of overall bias[b] −2 precision[c]
Suicide ideation	n on hormones = 42 n evaluated = 28 One prospective observational cohort study with mixed treatment (38 subjects with no pharmacological treatment)[16]	No change in suicide ideation	Cannot be assessed	−2 risk of overall bias[b] −2 precision[c]
Gender dysphoria	n on hormones = 145 n evaluated = 49 Two prospective observational cohort studies[15,16]	No change in gender dysphoria	Cannot be assessed	−2 risk of overall bias[b] −2 precision[c]
Depression	n on hormones = 97 n evaluated = 60 Two prospective observational cohort studies of which one included mixed treatment[14,16]	No change in depression	Cannot be assessed	−2 risk of overall bias[b] −2 precision[c]
Anxiety	n on hormones = 97 n evaluated = 60 Two prospective observational cohort studies[14,16]	No change in anxiety	Cannot be assessed	−2 risk of overall bias[b] −2 precision[c]
Cognition	n on hormones = 20 n evaluated = 20 One study[19]	No change in cognition compared with matched controls	Cannot be assessed	−2 risk of overall bias[b] −2 precision[c]
Quality of life	n on hormones = 98 n evaluated = 46 Two observational cohort studies, whereof one retrospective[16,17]	1. Improvement in quality of life most pronounced in subjects receiving puberty-blocking hormones, followed by gender-affirming hormone treatment[17] 2. Some improvement[16]	Cannot be assessed	−2 risk of overall bias[b] −2 precision[c]

Abbreviation: CGAS, Children's Global Assessment Scale.

[a]Starting at 4 for optimal studies in each study type.

[b]Selection of study participants is difficult to assess, analysis not based on stage in puberty development.

[c]Few study subjects in each study, heterogeneity in outcome and analyses.

could not investigate potential cognitive effects of hormone therapy.

3.5 | Bone health outcomes

Six longitudinal studies used dual-energy X-ray absorptiometry (DXA) scan technology to explore bone health before and again after some time with GnRHa treatment (Table 3). The second DXA scan usually coincided with CSHT initiation leading to different follow-up durations. The third DXA scan was performed after variable time with CSHT, performed with variable dosing and administration. The lumbar spine and hip were most often examined. One study investigated bone geometry.[20] Six studies were retrospective[21–26] and one study was prospective.[20] An additional study was cross-sectional where study participants in early puberty (Tanner stages 2–3) were examined only once, before the start of GnRHa therapy.[27]

Three studies reported a lower bone mineral density (BMD) in patients before or at start of GnRHa treatment compared with the general population of the same biological sex and age.[21,23,27] During GnRHa treatment, BMD estimated through area or volume, and expressed in z-scores increased less compared with general population reference values. However, the mean absolute BMD remained unchanged up to 2–3 years of GnRHa treatment.[20,23] The initiation of CSHT stimulated bone maturation and mineral accrual, increasing BMD.[21,22] After a median CSHT duration of 5.4 years in in female-to-male and 5.8 years in male-to-female, the lumbar spine mean areal BMD z-score was still significantly lower than at the start of GnRH therapy, while the other volume BMD and femoral neck estimates had normalised.[21] In another study, female-to-male receiving testosterone replacement therapy for 1–2 years had not regained their group mean BMD z-score registered at the start of GnRHa therapy.[24]

Bone geometry, estimated as subperiosteal width and endocortical diameter, was studied on DXA scans before start of GnRHa

TABLE 3 Summary of effects on bone development by puberty-blocking treatment (GnRHa) followed by CSHT in children with gender dysphoria.[20–25]

Outcome measures	Number of study participants, description of studies	Main Result	"Certainty of Evidence"	Deduction in GRADE[a]
Bone density during puberty-blocking hormonal treatment (g/cm^2, g/cm^3)	*n* on hormones = 363 *n* evaluated = 297 Five observational cohort studies (four retrospective and one prospective)[20–24]	Unchanged bone density (DXA measurement)	⊕⊕◯◯ Low certainty	−1 risk of overall bias[b] −1 precision
Bone density during puberty blocking hormonal treatment in relation to reference data in the literature (z-score)	*n* on hormones = 408 *n* evaluated = 292 Five observational cohort studies (four retrospective, and one prospective)[21–25]	Decreased increase in bone density over time	⊕⊕◯◯ Low certainty	−1 risk of overall bias[b] −1 precision
Bone density after 1–3 years (up to 22 years of age) of CSHT, which had been preceded by puberty-blocking hormonal treatment in relation to reference data in the literature	*n* on hormones = 268 *n* evaluated = 165 Three observational cohort studies (two retrospective and one prospective)[21,24,25]	After group median five years with CSHT, bone density recovered in hip but not in lumbar spine compared to data at start of treatment (z-score)	⊕⊕◯◯ Low certainty	−1 risk of overall bias[b] −1 precision

Abbreviations: CSHT, Cross-sex hormone treatment; DXA, Dual-Energy X-ray Absorptiometry.

[a]Starting at 4 for optimal studies in each study type.

[b]Analysis not based on stage in puberty development.

treatment and after at least two years on CSHT and compared with reference values of the general population: the bone geometry resembled the reference curve for the experienced sex only when GnRHa was started during early puberty. Bone geometry estimates in those who started GnRHa treatment during mid and late puberty remained within the reference curve of the biological sex.[26]

3.6 | Body composition and metabolic markers

GnRHa treatment effectively reduced endogenous sex hormone serum levels (Table 4). DXA scans after 1 year of GnRHa treatment revealed increased fat mass and reduced lean body mass.[28] Longitudinal growth depends on bone maturity (bone age) of those in the study group. Ongoing pubertal growth spurt will be arrested when GnRHa therapy is started, reducing the growth velocity to the prepubertal rate.[29]

Nokoff et al studied body composition and insulin sensitivity during 1 year of GnRHa therapy.[30] In addition to body composition, metabolic effects as insulin sensitivity during CSHT, and changes in blood pressure during testosterone therapy were examined.[31–33] Of these studies, three originated from Amsterdam.[29,30,33] The Amsterdam studies included observations during GnRHa therapy,[26] 1 year after starting CSHT,[32] as well as after a group median >5 years with CSHT in a cohort of 22-year-old adolescents.[31,33] The studies from Amsterdam were generally larger than the other studies. CSHT changed body composition towards the affirmed sex.[31,32] Obesity (defined as BMI >30 at age 22 years) was more prevalent in the transgender population[33] (Table 4).

3.7 | CSHT in children without prior GnRHa treatment

We were able to identify three studies of low-to-moderate bias examining CSHT in children without prior GnRHa treatment.[13,34,35] All were retrospective longitudinal studies. Because the number of study participants was small, studies were deemed to have low external validity, and because the studies examined different outcomes (e.g., lipid serum levels, Hb, blood pressure, metrorrhagia), it was not possible to draw any overall conclusions from these studies. Although the Mullins et al. paper[13] included several individuals at elevated risk of arterial or venous thrombosis, no cases of thrombosis were reported.

4 | DISCUSSION

We performed an extensive literature search to examine psychosocial and cognitive outcomes as well as metabolic and bone health in children with gender dysphoria taking hormone therapy. No randomised controlled trials were found, but we could identify 24 relevant observational studies. However, these were limited by

TABLE 4 Summary of findings of puberty-blocking (GnRHa) hormone treatment on anthropometric measures, body composition, and metabolism in children with gender dysphoria.[28–33]

Outcome measures	Number of study participants, description of studies	Main result	"Certainty of Evidence"	Deduction in GRADE[a]
Anthropometric measures	*n* on hormones = 192 *n* evaluated = 192 One retrospective observational cohort study[31]	Increased weight and body mass index	Cannot be assessed	−2 risk for overall bias[b] −1 precision[c] −1 indirectness[d]
Body composition	*n* on hormones = 325 *n* evaluated = 286 Two prospective observational cohort studies and one controlled cross-sectional study[28,30,31]	Decreased lean body mass	Cannot be assessed	−2 risk for overall bias[b] −1 precision[c] −1 indirectness[d]
Metabolic measures	*n* on hormones = 209 *n* evaluated = 209 One retrospective observational cohort study and one controlled cross-sectional study[30,32]	No change in serum lipids or blood pressure Increased insulin level in MtF Decreased insulin sensitivity	Cannot be assessed	−2 risk for overall bias[b] −1 precision[c] −1 indirectness[d]
Blood pressure	*n* on hormones = 15 *n* evaluated = 15 One retrospective observational cohort study[33]	Change in blood pressure	Cannot be assessed	−2 risk for overall bias[b] −1 precision[c] −1 indirectness[d]
Growth (cm/year)	*n* on hormones = 55 *n* evaluated = 55 One prospective multicentre observational GnRHa treatment cohort study[29]	Reduced growth velocity	Cannot be assessed	−2 risk for overall bias[b] −1 precision[c] −1 indirectness[d]

[a]Starting at 4 for optimal studies in each study type.
[b]Selection of study participants is difficult to assess. Analysis not based on stage in puberty development.
[c]Few study subjects in each study, hence there is heterogeneity in outcome and analyses.
[d]Single study. In this context, 'indirectness' is similar to 'external validity'.

methodological weaknesses, for instance lack of or inappropriate control group, lack of intra-individual analyses, high attrition rates that precluded conclusion to be drawn. The exception being that children with gender dysphoria often had lower group mean values for BMD already prior to GnRHa treatment, and that GnRHa treatment delays the physiologically occurring BMD gain during pubertal sex hormone stimulation. However, this GnRHa-induced delay in BMD gain is almost fully compensated for by later ensuing CSHT. Although study participants were followed up to 22 years of age, the observed remaining deficit may depend on the limited study group size or on too short observation time.[21]

Our review highlights several specific knowledge gaps in gender dysphoria that are important to bridge not least given the recent increased incidence in many countries.[6,7] First, randomised controlled trials are lacking in gender dysphoria research. We call for such studies, which may be the only way to address biases that we have noted in the field. Given the current lack of evidence for hormonal therapy improving gender dysphoria, another ethically feasible option would be to randomise individuals to hormone therapy with all study participants, independent of intervention status, receiving psychological and psychosocial support. However, controlled trials do not necessarily require placebo treatment, but could for example build on the date or time of starting hormonal therapy to generate comparison groups. However, it should also be noted that this is a highly vulnerable population.

A second limitation concerns the statistical management of data. In the reviewed studies, observational data have frequently been analysed at a group level where intra-individual changes would have been more appropriate. Intra-individual analyses would allow for a better understanding of how subgroups of individuals respond (both positively and negatively) to hormone therapy. Group-level analyses are sensitive to selection bias because of high drop-out rates: The group studied at the end of the study is a selection of the group studied at baseline, which increases indirectness (reduces external validity). Moreover, it is important to analyse the distribution of individual data to be able to identify outliers who may be at risk for severe consequences of treatment.

Third, many studies only present data on chronological age but fail to account for puberty stage and biological age. This is a concern because the main purpose of GnRHa treatment is to suppress puberty and, with that, biological ageing.

Fourth, long-term studies are lacking. The duration of GnRHa treatment and CSHT was rarely >4 years. The absence of long-term studies is worrying because many individuals start treatment as minors (<18 years) and CSHT is lifelong. Fifth, individuals who stop GnRHa treatment before the start of CSHT need to be described and followed up. Sixth, some of the findings underlying this review are old, and studies reflecting the changing demographics of individuals seeking care for gender dysphoria are warranted.

TABLE 5 The GEnder Dysphoria HORmone treatment (GENDHOR) checklist.

	Recommendations
Aim	Describe the aim of the study
Study participants:	
Cases/exposed	Define gender dysphoria in your study, including the assessment tools used. Define eligibility criteria for your study (including chronological age, bone age or puberty stage, according to Tanner or Prader (when study concerns adolescents), biological sex, perceived gender identity, psychiatric and somatic comorbidities, medications at baseline). List exclusion criteria (diagnoses). List ages of participants at the start of each treatment (including absolute age ranges).
Comparators/ unexposed	Clarify how controls were selected (were controls recruited from the general population?) or whether national/ regional reference data (for instance, Z-scores) were used instead of individual controls.
Study design	Describe the study design: Cross-sectional, retrospective, prospective; case–control (and if nested), cohort study, randomised clinical trial.
Setting	Describe the setting of the study. Were study participants included at a tertiary centre or from the general population? Describe the catchment area/population of participating centres.
Intervention	Hormone treatment Describe whether GnRHa, anti-androgens, CSHT, or a combination was used. List generic names, mode of administration, and dosages of all treatments. Specify the treatment duration of each treatment. If hormone serum concentrations are studied, include the standard procedure for the timing of blood samples to hormone intake. If patients undergo surgery, clarify the type of surgery and number of participants undergoing each surgical procedure (gonadectomy, mastectomy, laryngeal surgery, vaginoplasty/phalloplasty, etc.). Clarify if any participant received psychiatric counselling before, or during the study, including total duration and frequency of counselling.
Variables	Define each variable (including co-variates) and its source. If possible, mention any effort to validate the variables.
Data measurement	Clarify who collected the data on study participants. Present time between first and second measurements if your study is longitudinal and includes "before-after" measurements in relation to the intervention. Mention if study participants had previously been included in other studies with a different aim or examining other outcomes.
Blinding	Describe if the data collectors were blinded to participant status/treatment or not.
Loss to follow-up	Indicate the number of participants discontinuing GnRHa/ CSHT and the reason(s) for discontinuation, including no longer wish to pursue gender reassignment treatment. Describe loss to follow-up/missing data
Statistical methods	Describe statistics according to a relevant checklist. Consider when applicable: Intra-individual changes (mean, SD, median, range) vs. between-group differences.
Descriptive data	In addition to usual demographic, clinical, social/socioeconomic information, report body mass index (BMI), smoking, use of oral contraceptives (type) or other hormonal treatment, puberty stage. Report any psychiatric illness at baseline, as well as the use of psychotropic medication. Describe other comorbidities, including disorders that could be considered contraindications for either hormone treatment or surgery. Specify follow-up time (median, mean) since the start of the intervention and since start of hormone treatment (define intervention start).
Outcome data	Specify main outcome of the study. Indicate all secondary outcomes, including adverse events.
Adverse events/ complications	Describe all adverse events.
Main results	Present absolute numbers. Calculate absolute and relative risks/Intraindividual effects/change and group mean/ median. Present incidence data. Describe any adjustment for potential confounders.
Limitations	Discuss limitations of your study, including limitations of the measurements used (e.g., DXA) and sources of potential bias or imprecision.
Generalisability/ external validity	Can data be generalised to individuals with gender dysphoria outside your study centre and the study country?
Conflict of interest	Report any conflict of interest.

Note: Based on our literature review, we created a *GEnder Dysphoria HORmone treatment checklist* (GENDHOR).

This list consists of recommendations that researchers may consider when planning a study of gender dysphoria, whether observational or interventional.

Abbreviations: CSHT, Cross-sex hormone treatment; DXA, Dual-Energy X-ray Absorptiometry; GnRHa, Gonadotropin-releasing hormone agonist (analogues).

Finally, we could not evaluate the frequency of individuals who drop out from GnRHa treatment and no longer wish to continue with gender transition. However, a follow up study was published after our literature search.[36] Of 720 children (31% born male and 69% born female) who started GnRHa treatment in adolescence, 98% continued to use hormone treatment into adulthood, which suggests that children generally continue with gender transition once they have started GnRHa treatment. We know from internet-based surveys that detransitioning exists,[37] but such studies cannot provide reliable estimates of detransitioning frequency because of selection bias. Studies that closely follow individuals who start GnRHa therapy and/or CSHT until at least age 30 are urgently needed. We also acknowledge there are other potential side effects from GnRHa therapy or CSHT that were not included in our review such as alopecia and abscesses from injections.[38]

Due to limitations in reporting of data, previous published studies in this field repeatedly contain insufficient details on drug administration and dosages, treatment duration, and the type of surgery performed. Some of these limitations will be partly remedied by the introduction of the new ICD version 11, and the Utrecht criteria,[39] but the field also urgently needs high quality longitudinal studies that not only assess medical outcomes but also those outcomes that matter most for affected individuals. Building on the identified limitations in previous research, we compiled a checklist to improve gender dysphoria research ("GENDHOR", Table 5). The aim of this checklist is not to replace existing research guidelines, but using it together with existing guidelines might support researchers and peer reviewers, and ultimately benefit patients and their families.

Last, there have been studies in this field published after the date of our literature search (9 November 2021). These have not been added to this study in order to not depart from the systematic approach. We nevertheless wish to comment on some of the publications. First, the National Institute for Health and Care Excellence in England (NICE) conducted evidence reviews of GnRHa[40] as well as CSHT[41] for children with gender dysphoria, which were independent from our work. The conclusions generally align with our findings. Second, Chien et al.[42] recently published a prospective study of psychosocial functioning during 2 years after initiation of CSHT in youths (12–20 years of age) with gender dysphoria. Of 315 participants, 162 completed that study. Life satisfaction increased, and depression and anxiety scores decreased, among biological females but not biological males. The strongest finding was a moderately improved appearance congruence. No information on concomitant psychological or psychopharmacological therapy was provided.

5 | CONCLUSION

This systematic review of almost 10000 screened abstracts suggests that long-term effects of hormone therapy on psychosocial and somatic health are unknown, except that GnRHa treatment seems to delay bone maturation and gain in bone mineral density.

AUTHOR CONTRIBUTIONS

Study concept and design: All authors. Acquisition of data: Malin Höistad, Jan Adolfsson. Drafting of the manuscript: All authors. Interpretation of data and critical revision of the manuscript for important intellectual content: All authors. Administrative, technical, or material support: Jan Adolfsson, Malin Höistad. Funding acquisition: the Swedish agency for technology assessment and assessment for social services.

ACKNOWLEDGEMENTS

Information specialists Klas Moberg and Hanna Olofsson designed and performed the literature search.

FUNDING INFORMATION

This study received funding from the Swedish Agency for Health Technology Assessment and Assessment of Social Services.

CONFLICT OF INTEREST STATEMENT

JFL coordinated an unrelated study on behalf of the Swedish inflammatory bowel disease quality register (SWIBREG) that received funding from the Janssen Corporation. JFL has also received financial support from Merck Sharp & Dohme developing a paper reviewing national healthcare registers in China. JFL is currently discussing potential research collaboration with Takeda. ML has received lecture honoraria for Lundbeck pharmaceuticals and served as consultant for AstraZeneca. The other authors report no conflict of interest.

The data collection of this study was funded by the Swedish agency for technology assessment and assessment for social services. JA and MH are employees at this agency.

ORCID

Jonas F. Ludvigsson https://orcid.org/0000-0003-1024-5602
Jan Adolfsson https://orcid.org/0000-0003-2992-1869
Malin Höistad https://orcid.org/0000-0002-2312-3885
Per-Anders Rydelius https://orcid.org/0000-0002-1923-0282
Berit Kriström https://orcid.org/0000-0002-5456-2514
Mikael Landén https://orcid.org/0000-0002-4496-6451

REFERENCES

1. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic Criteria for Research. WHO; 1993.
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. American Psychiatric Association; 2013.
3. Coleman E, Bockting W, Botzer M, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. Int J Transgend. 2012;13(4):165-232. doi:10.1080/15532739.2011.700873
4. Cohen-Kettenis PT, van Goozen SH. Pubertal delay as an aid in diagnosis and treatment of a transsexual adolescent. Eur Child Adolesc Psychiatry. 1998;7(4):246-248. doi:10.1007/s007870050073
5. de Vries AL, Cohen-Kettenis PT. Clinical management of gender dysphoria in children and adolescents: the Dutch approach. J Homosex. 2012;59(3):301-320. doi:10.1080/00918369.2012.653300

6. Hisle-Gorman E, Schvey NA, Adirim TA, et al. Mental healthcare utilization of transgender youth before and after affirming treatment. J Sex Med. 2021;18(8):1444-1454. doi:10.1016/j.jsxm.2021.05.014
7. Landén M. Dramatic increase in adolescent gender dysphoria requires careful consideration. Lakartidningen. 2019;116:FSMH.
8. Thompson L, Sarovic D, Wilson P, Sämfjord A, Gillberg C. A PRISMA systematic review of adolescent gender dysphoria literature: (1) Epidemiology. PLOS Glob Public Health. 2022;2(3):e0000245. doi:10.1371/journal.pgph.0000245
9. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4:1. doi:10.1186/2046-4053-4-1
10. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919. doi:10.1136/bmj.i4919
11. SBU. Utvärdering av metoder i hälso- och sjukvården och insatser i socialtjänsten: en metodbok. Stockholm. Statens beredning för medicinsk och social utvärdering (SBU); 2020. Available from: https://www.sbu.se/metodbok
12. Guyatt GH, Oxman AD, Schunemann HJ, et al. GRADE guidelines: a new series of articles in the journal of clinical epidemiology. J Clin Epidemiol. 2011;64(4):380-382. doi:10.1016/j.jclinepi.2010.09.011
13. Mullins ES, Geer R, Metcalf M, et al. Thrombosis risk in transgender adolescents receiving gender-affirming hormone therapy. Pediatrics. 2021;147(4):e2020023549. doi:10.1542/peds.2020-023549
14. de Vries AL, McGuire JK, Steensma TD, et al. Young adult psychological outcome after puberty suppression and gender reassignment. Pediatrics. 2014;134(4):696-704. doi:10.1542/peds.2013-2958
15. Costa R, Dunsford M, Skagerberg E, Holt V, Carmichael P, Colizzi M. Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. J Sex Med. 2015;12(11):2206-2214. doi:10.1111/jsm.13034
16. Carmichael P, Butler G, Masic U, et al. Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK. PLoS One. 2021;16(2):e0243894. doi:10.1371/journal.pone.0243894
17. Becker-Hebly I, Fahrenkrug S, Campion F, Richter-Appelt H, Schulte-Markwort M, Barkmann C. Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: a descriptive study from the Hamburg gender identity service. Eur Child Adolesc Psychiatry. 2021;30(11):1755-1767. doi:10.1007/s00787-020-01640-2
18. Cantu AL, Moyer DN, Connelly KJ, Holley AL. Changes in anxiety and depression from intake to first follow-up among transgender youth in a pediatric endocrinology clinic. Transgend Health. 2020;5(3):196-200. doi:10.1089/trgh.2019.0077
19. Staphorsius AS, Kreukels BP, Cohen-Kettenis PT, et al. Puberty suppression and executive functioning: an fMRI-study in adolescents with gender dysphoria. Psychoneuroendocrinology. 2015;56:190-199. doi:10.1016/j.psyneuen.2015.03.007
20. Schagen SEE, Wouters FM, Cohen-Kettenis PT, Gooren LJ, Hannema SE. Bone development in transgender adolescents treated with GnRH analogues and subsequent gender-affirming hormones. J Clin Endocrinol Metab. 2020;105(12):e4252-e4263. doi:10.1210/clinem/dgaa604
21. Klink D, Caris M, Heijboer A, van Trotsenburg M, Rotteveel J. Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. J Clin Endocrinol Metab. 2015;100(2):E270-E275. doi:10.1210/jc.2014-2439
22. Vlot MC, Klink DT, den Heijer M, Blankenstein MA, Rotteveel J, Heijboer AC. Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone. 2017;95:11-19. doi:10.1016/j.bone.2016.11.008
23. Joseph T, Ting J, Butler G. The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort. J Pediatr Endocrinol Metab. 2019;32(10):1077-1081. doi:10.1515/jpem-2019-0046
24. Stoffers IE, de Vries MC, Hannema SE. Physical changes, laboratory parameters, and bone mineral density during testosterone treatment in adolescents with gender dysphoria. J Sex Med. 2019;16(9):1459-1468. doi:10.1016/j.jsxm.2019.06.014
25. Navabi B, Tang K, Khatchadourian K, Lawson ML. Pubertal suppression, bone mass, and body composition in youth with gender dysphoria. Pediatrics. 2021;148(4):e2020039339. doi:10.1542/peds.2020-039339
26. van der Loos MA, Hellinga I, Vlot MC, et al. Development of hip bone geometry during gender-affirming hormone therapy in transgender adolescents resembles that of the experienced gender when pubertal suspension is started in early puberty. J Bone Miner Res. 2021;36(5):931-941. doi:10.1002/jbmr.4262
27. Lee JY, Finlayson C, Olson-Kennedy J, et al. Low bone mineral density in early pubertal transgender/gender diverse youth: findings from the trans youth care study. J Endocr Soc. 2020;4(9):bvaa065. doi:10.1210/jendso/bvaa065
28. Schagen SE, Cohen-Kettenis PT, Delemarre-van de Waal HA, et al. Efficacy and safety of gonadotropin-releasing hormone agonist treatment to suppress puberty in gender dysphoric adolescents. J Sex Med. 2016;13(7):1125-1132. doi:10.1016/j.jsxm.2016.05.004
29. Schulmeister C, Millington K, Kaufman M, et al. Growth in transgender/gender-diverse youth in the first year of treatment with gonadotropin-releasing hormone agonists. J Adolesc Health. 2022;70(1):108-113. doi:10.1016/j.jadohealth.2021.06.022
30. Nokoff NJ, Scarbro SL, Moreau KL, et al. Body composition and markers of cardiometabolic health in transgender youth compared with cisgender youth. J Clin Endocrinol Metab. 2020;105(3):e704-e714. doi:10.1210/clinem/dgz029
31. Klaver M, de Mutsert R, Wiepjes CM, et al. Early hormonal treatment affects body composition and body shape in young transgender adolescents. J Sex Med. 2018;15(2):251-260. doi:10.1016/j.jsxm.2017.12.009
32. Klaver M, de Mutsert R, van der Loos M, et al. Hormonal treatment and cardiovascular risk profile in transgender adolescents. Pediatrics. 2020;145(3):e20190741. doi:10.1542/peds.2019-0741
33. Perl L, Segev-Becker A, Israeli G, Elkon-Tamir E, Oren A. Blood pressure dynamics after pubertal suppression with gonadotropin-releasing hormone analogs followed by testosterone treatment in transgender male adolescents: a pilot study. LGBT Health. 2020;7(6):340-344. doi:10.1089/lgbt.2020.0026
34. Tack LJ, Craen M, Dhondt K, et al. Consecutive lynestrenol and cross-sex hormone treatment in biological female adolescents with gender dysphoria: a retrospective analysis. Biol Sex Differ. 2016;7:14. doi:10.1186/s13293-016-0067-9
35. Jarin J, Pine-Twaddell E, Trotman G, et al. Cross-sex hormones and metabolic parameters in adolescents with gender dysphoria. Pediatrics. 2017;139(5):e20163173. doi:10.1542/peds.2016-3173
36. van der Loos M, Hannema SE, Klink DT, et al. Continuation of gender-affirming hormones in transgender people starting puberty suppression in adolescence: a cohort study in The Netherlands. Lancet Child Adolesc Health. 2022;6:869-875. doi:10.1016/S2352-4642(22)00254-1
37. Littman L. Individuals treated for gender dysphoria with medical and/or surgical transition who subsequently detransitioned: a survey of 100 detransitioners. Arch Sex Behav. 2021;50(8):3353-3369. doi:10.1007/s10508-021-02163-w

38. Khatchadourian K, Amed S, Metzger DL. Clinical management of youth with gender dysphoria in Vancouver. J Pediatr. 2014;164(4):906-911. doi:10.1016/j.jpeds.2013.10.068
39. McGuire JK, Berg D, Catalpa JM, et al. Utrecht gender dysphoria scale - gender Spectrum (UGDS-GS): construct validity among transgender, nonbinary, and LGBQ samples. Int J Transgend Health. 2020;21(2):194-208. doi:10.1080/26895269.2020.1723460
40. Evidence review: gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria (NICE. National Institute for Health and Care Excellence). 2020 https://cassindependent-reviewuk/nice-evidence-reviews/
41. Evidence review: gender-affirming hormones for children and adolescents with gender dysphoria (NICE. National Institute for Health and Care Excellence). 2020 https://cass.independent-review.uk/nice-evidence-reviews/
42. Chen D, Berona J, Chan YM, et al. Psychosocial functioning in transgender youth after 2years of hormones. N Engl J Med. 2023;388(3):240-250. doi:10.1056/NEJMoa2206297

How to cite this article: Ludvigsson JF, Adolfsson J, Höistad M, Rydelius P-A, Kriström B, Landén M. A systematic review of hormone treatment for children with gender dysphoria and recommendations for research. Acta Paediatr. 2023;112:2279-2292. https://doi.org/10.1111/apa.16791

EDITORIAL

Editorial: Dialectical Behavior Therapy: More Is Not Always Better When Different Is Required

Aron Janssen, MD, and John T. Walkup, MD

The study by Berk *et al.*[1] highlights potential trajectories of response and nonresponse to dialectical behavior therapy (DBT) as compared to individual and group supportive therapy (IGST) for teens with repeated self-harm and suicidal ideation. The authors also posit a testable function to predict responsiveness vs nonresponsiveness and provide critical guidance about when to reassess nonresponders and alter treatment. This is the fourth major article from a large federally funded, randomized controlled trial. Previous publications have highlighted superiority of DBT over IGST,[2] reported the moderating factors of treatment outcomes,[3] and explored the mechanism of effectiveness for DBT in the treatment of suicidal ideation and self-harm.[4] These articles provide useful information given the rising rates of suicidal ideation and suicide attempts among youth[5] and recent research suggesting the powerful role of social media in supporting contagion of suicidal behavior among youth.

Rising rates of youth suicide were a public health concern even before the coronavirus disease 2019 (COVID-19) pandemic. Rates of suicide increased 3% from 2007 to 2014 and increased by 10% from 2014 to 2017.[6] It remains the second leading cause of death among teens, and for the first time suicide has appeared among the top 10 causes of death for children ages 5–9.[6,7]

Contagion of suicidal behavioral among youth and their high level of engagement with social media are particularly concerning.[8] Never before have we seen such overexposure to peers' suicidal behavior paired with the degradation of the usual prosocial community contexts that can promote healthy coping. Children now have in their pockets access to communities and information that encourage exploration of suicide and teach children how to access more lethal means online. Given that previous suicide attempts and exposure to suicide attempts among one's peer group increase risk for future suicide attempts, it is alarming to imagine the potential lifetime of increased risk for this cohort of youth.

Our mental health care system for youth was broken before the COVID-19 pandemic with limited financial support for mental health care resulting in too few providers to address the mental health needs of the upward 20% of affected youth. Perhaps of even greater concern is the limited uptake of evidenced-based interventions among existing providers. The mental health pandemic has placed this fundamental moral and ethical challenge front and center for our society.

To date, evidence-based interventions to address suicidal behavior in youth are limited. DBT was first developed to treat adults with personality disorders characterized by emotion dysregulation and complicated by suicidal behavior and self-injury. Recent adaptations of DBT for adolescents and children hold promise to intervene much earlier in life, mitigate current impairment and distress, and prevent the evolution of emotion dysregulation and its sequelae into adulthood.

DBT as implemented in Berk *et al.*[1] consists of multiple components, including individual, group, and family work as well as access to 24/7 phone coaching. For suicidal ideation, this study shows a mixed picture for the effectiveness of DBT as compared to IGST. There was clear superiority demonstrated for DBT over IGST for self-harm, but for not for suicidal ideation. Of participants, 63% demonstrated responsiveness to care and a reduction in suicidal ideation, meaning that 37% did not respond despite the intensity of both interventions. It should be noted that any study comparing interventions with a focus on suicidal behavior is compelled to ensure the safety of participants in both intervention groups. Thus, the lack of between-group differences and the overall response rate suggest that both interventions worked to ensure the safety of participants. In a secondary analysis, the study team differentiated participants into those who showed improvement in suicidal ideation or self-harm (responders) and those described as "total nonresponders." While it is vital to understand what interventions work most effectively, it is just as important to understand for whom

these interventions work and for whom these interventions are ineffective. As DBT (and IGST) is an intensive intervention that is expensive to implement, knowing what works for whom allows for a more effective match between patient need and intervention and improvement of access to care for those likeliest to respond. In their analysis, the study team noted several factors that predicted nonresponse, including White race, nightmares, higher parent-reported sleep problems, higher numbers of past self-harm episodes in the 3 months before baseline, higher internalizing symptoms on the Youth Self-Report and/or Child Behavior Checklist, higher youth-reported emotion dysregulation on the Difficulties in Emotion Regulation Scale (DERS), and higher suicidal ideation.[1] Notably, the participants who responded to care and the participants who were nonresponders separated by the third month of treatment. The authors are wise to note that a reassessment of the treatment plan is warranted for nonresponse at 3 months, and one large potential benefit of this study is to encourage providers to reassess their care earlier in treatment than may typically be done.

While the study team identified baseline characteristics predicting response, they have not been able to identify exactly why such characteristics (short of symptom severity) are associated with lack of response. It will be important to replicate these findings across different samples that are more balanced across gender categories (including more males), include more information from the course of treatment that could help clarify appropriate next steps in care and further differentiate who will and will not respond, and, vitally, identify what must be changed or adapted for the nonresponders. For example, we would want to understand specific diagnoses that may predict responsiveness vs nonresponsiveness to care, sexual and gender minority status, exposure to trauma, and use of medication. The authors note that higher internalizing symptoms predict nonresponse, but we are left to wonder if these patients were receiving appropriate pharmacologic interventions for depression or anxiety. Although the authors did stratify based on use of psychiatric medication, there was no indication of what the medication was, if it changed over time, or if there were differences in medication class across the study groups. Future analyses done by this research group or by others will continue to help systems of care and individual providers tailor interventions more effectively.

Berk *et al.*[1] make it very clear that DBT and IGST, when provided with fidelity, are highly effective interventions that treat either self-harm or suicidal ideation, with 87% of participants responding to at least 1 domain. Yet, little is known about why youth do not respond and how widely generalizable these results may be in a more diverse population. We would also call for more research into the practical application of these interventions, given the likelihood that most available providers of DBT and IGST are likely not providing the same intensity or quality of care as that of the research team. Furthermore, there are practical limitations accessing DBT for youth with Medicaid, and skilled DBT providers likely limit their caseload to provide DBT at this level and further strain the mental health care system. However, these limitations are in some ways the biggest strength of the article: what Berk *et al.*[1] make clear is that we must not shy away from understanding the limitations of a particular intervention, that we should advocate as fiercely as possible for getting all youth access to the best care, and that we should reassess the effectiveness of that care much sooner in the process and change the treatment plan accordingly. More is not always better when different is required.

Accepted May 27, 2022.

Drs. Janssen and Walkup are with Ann and Robert H. Lurie Children's Hospital of Chicago, Illinois, and Northwestern University Feinberg School of Medicine, Chicago, Illinois.

The authors have reported no funding for this work.

Author Contributions:
Conceptualization: Janssen, Walkup.
Writing — original draft: Janssen, Walkup.
Writing — review and editing: Janssen, Walkup.

Disclosure: Dr. Walkup has served on the advisory board and Speaker's Bureau of the Tourette Association of America and has received royalties from Oxford Press, Guilford Press, and Wolters Kluwer. Dr. Janssen has reported no biomedical financial interests or potential conflicts of interest.

Correspondence to John T. Walkup, MD, Ann and Robert H. Lurie Children's Hospital of Chicago, 225 E. Chicago Avenue, Box 10, Chicago, IL 60611; e-mail: JWalkup@luriechildrens.org)

0890-8567/$36.00/©2022 American Academy of Child and Adolescent Psychiatry

https://doi.org/10.1016/j.jaac.2022.05.006

REFERENCES

1. Berk MS, Gallop R, Asarnow JR, *et al.* Trajectories of treatment response and nonresponse in youth at high risk for suicide. J Am Acad Child Adolesc Psychiatry. 2022;61:1119-1130. https://doi.org/10.1016/j.jaac.2022.01.010
2. McCauley E, Berk MS, Asarnow JR, *et al.* Efficacy of dialectical behavior therapy for adolescents at high risk for suicide: A randomized clinical trial. JAMA Psychiatry. 2018;75:777-785. https://doi.org/10.1001/jamapsychiatry.2018.1109
3. Adrian M, McCauley E, Berk MS, *et al.* Predictors and moderators of recurring self-harm in adolescents participating in a comparative treatment trial of psychological interventions. J Child Psychol Psychiatry. 2019;60:1125-1132. https://doi.org/10.1111/jcpp.13099
4. Asarnow JR, Berk MS, Bridics J, *et al.* Dialectical behavior therapy for suicidal self-harming youth: Emotion regulation, mechanisms, and mediation. J Am Acad Child Adolesc Psychiatry. 2021;60:1105-1115.

5. Hill RM, Rufino K, Kurian S, Saxena J, Saxena K, Williams L. Suicide ideation and attempts in a pediatric emergency department before and during COVID-19. Pediatrics. 2021;147:e2020029280. https://doi.org/10.1542/peds.2020-029280
6. Centers for Disease Control and Prevention. About underlying cause of death, 1999-2019. Accessed July 1, 2022; https://wonder.cdc.gov/ucd-icd10.html
7. Schlagbaum P, Tissue T, Sheftall, *et al*. The impact of peer influencing on adolescent suicidal ideation and suicide attempts. J Psychiatr Res. 2021;140:529-532. https://doi.org/10.1016/j.jpsychires.2021.06.027
8. Sedgwick R, Epstein S, Dutta R, Ougrin D. Social media, internet use and suicide attempts in adolescents. Curr Opin Psychiatry. 2019 Nov;32(6):534-541. https://doi.org/10.1097/YCO.0000000000000547

All statements expressed in this column are those of the authors and do not reflect the opinions of the *Journal of the American Academy of Child and Adolescent Psychiatry*. See the Guide for Authors for information about the preparation and submission of Editorials.

IN THE UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF ALABAMA
NORTHERN DIVISION

BRIANNA BOE, individually and on behalf of her minor son, MICHAEL BOE; *et al.*,

Plaintiffs,

and

UNITED STATES OF AMERICA,

Plaintiff-Intervenor,

v.

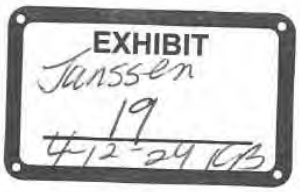
STEVE MARSHALL, in his official capacity as Attorney General of the State of Alabama; *et al.*,

Defendants.

Case No. 2:22-cv-00184-LCB-CWB

Honorable Liles C. Burke

SUPPLEMENTAL EXPERT REPORT OF
ARON JANSSEN, MD



1. My name is Aron Janssen, M.D. I am a board-certified child and adolescent psychiatrist. I specialize in the treatment of gender dysphoria in children and adolescents. On February 10, 2023, I provided a report on the standards of care for treating individuals diagnosed with gender dysphoria. I have been asked to supplement that report based on my review of the following supplemental expert reports:

a. Supplemental Expert Report of James Cantor, Ph.D. (Feb. 2, 2024)

b. Supplemental Expert Report of Michael K. Laidlaw, M.D. (Feb. 2, 2024)

2. The articles Cantor and Laidlaw rely on are largely systematic reviews of data that have been available for many years. Cantor and Laidlaw use this older data to draw unfounded conclusions they wrongly suggest weigh against providing treatment options. Nothing in the articles they cite rebuts my earlier opinion regarding the efficacy of gender transition treatments when medically indicated.

3. Cantor and Laidlaw also levy unsupported criticisms of the WPATH process for generating the Standards of Care 8 ("SOC 8"). Rather than diminishing the strength of the Delphi-approved recommendations of the SOC 8, the materials Defendants point to show the rigorous and informed debate behind the development of SOC 8. The recommendations in the SOC 8 went through a standard Delphi process that included consensus votes. Some of the recommendations were supported on the first consensus consideration. Others were discussed and debated

over time and then refined. That is part of scientific and rigorous debate and fully consistent with the Delphi process, which is designed to foster robust discussion and airing of views. Disagreements among scientists and professionals reflect the robustness of the process.

4. Finally, Defendants' experts agree that gender dysphoria is a real, serious and debilitating medical condition for adolescents. The scientific literature shows that gender transition medications can diminish or alleviate gender dysphoria, and the State's experts can point to no evidence that shows any effective alternative treatment for gender dysphoria in adolescents.

I. CANTOR

5. Cantor bases his opinion on isolated statements from systematic reviews suggesting a need for more evidence to strengthen what we know about the efficacy of established medical care for adolescents with gender dysphoria. There is virtually no area of medical treatment or research, if any, in which such systematic reviews would not include similar statements noting the need for further research; in fact, one of the main purposes of systematic reviews is to flag such needs. Cantor's opinion is largely based on this fundamental misreading of these isolated statements, not on any meaningful review of the substantial body of research on this area of medical care and not on any day-to-day experiences in treating transgender patients or on his clinical observations.

6. I would agree with Cantor that there is no one assessment tool that is guaranteed to capture internal signals that can sometimes be misread as related to gender dysphoria. Accordingly, in actual practice, assessment tools can be a component of assessment but are not the entirety of it. For example, the standard assessment for gender dysphoria includes, and has always included, an assessment of the social milieus with which the individual interacts. One must understand the social context in which an individual patient lives in order to make a diagnosis and be able to identify a patient's identity as being stable over time. This is a core tenet of gender-affirming care.

A. Cantor's Commentary on Studies

1. Morandini

7. Morandini et al. (2023) use a cross-sectional sample of children and adolescents ages 4-17 to assess the mental health effects of social gender transition. The authors examine the mental health of gender dysphoric children and adolescents referred to the Gender Identity Development Service (GIDS) using clinician ratings of mental health, comparing the mental health of those who had socially transitioned to those who had not. (p. 1048). Overall, the authors generally found that children and adolescents who had socially transitioned and those who had not did not differ

with regard to narrowly defined mental health status in the short term.[1] (pp. 1052-53). In the study, the authors used anxiety, depression and suicidality measures to represent the broad category of mental health. However, specific measures of intensity of gender dysphoria and the impact on quality of life were not included in this analysis, and it is the distress from the core symptoms of Gender Dysphoria that social transition, along with the other aspects of gender affirming medical care aim to target.

8. By virtue of the cross-section research design—and by the authors' own admission—the study captures some aspects of mental health at one specific point in time: upon the children and adolescents' first contact with GIDS, at which time clinicians assessed mental health. Because the study focuses on a snapshot in time, it does not support a conclusion that there are no mental health benefits of social transition over a longer time horizon. While the study adds to the field in important ways, it does so modestly, and the authors caution:

> "Our data suggest that social gender transition *may not render immediate and dramatic alleviation of mental health difficulties* for all or most children/adolescents suffering with gender dysphoria. If it did, we would expect to have found some lower prevalence of anxiety or depression in our socially transitioned group. Perhaps our study hints that *social gender transition alone, at least in the short term, is no panacea to mental health struggles of young people with gender dysphoria* and that clinicians and parents should not expect immediate

[1] They did find that mood disorders were more common among children and adolescents assigned male at birth who did not transition but express that the finding may be spurious.

> symptom alleviation specific to gender dysphoria or related to mental health more generally." (p. 1057).

Furthermore, the study is specific to social transitioning and does not support any negative inferences with regard to the efficacy of gender transition medications. If patients have an adjustment disorder in which their depression or anxiety or suicidality is a direct result of gender dysphoria, we often see individualized improvements from social transition. But the long-term positive effects of social transition alone, without gender transition medications, are often limited.

This study demonstrated merely that social transition in and of itself is not a cure-all for gender dysphoria in all cases and may not improve co-occurring psychiatric conditions. That said, as the authors acknowledge, "[i]t is possible that although our socially transitioned patients did not demonstrate superior well-being compared to their non-transitioned counterparts, they were nevertheless functioning better (either in terms of mood/anxiety or gender dysphoria severity, or both) than their own prior functioning pre-social transition." (p. 1057). This is an important caveat as this study provides no way to assess whether social transition improved the mental health of these patients relative to their mental health before socially transitioning.

9. In my opinion, Dr. Cantor overstates the conclusions from this study, and does so in ways the authors themselves expressly caution against. Some patients with

gender dysphoria also have independent mental health diagnoses such as depression or anxiety. Social transition is treatment for gender dysphoria. It may not alleviate other independent mental health conditions from which a patient suffers. That does not mean that it is not effective at alleviating gender dysphoria.

10. It is also unrealistic to conduct a randomized control trial for social transition as a patient cannot be blinded from the treatment.

2. Glintborg et al. 2023 and Kaltiala (2023)

11. Glintborg et al. (2023) investigates mental health outcomes for a large sample of Danish transgender individuals after initiation of gender-affirming care. The authors examined longitudinal outcomes for mental and behavioral disorders and prescription of psychopharmacological agents in almost 4,000 individuals following their initial diagnosis of gender incongruence. (p. 338-39). They found that the risk for mental and behavioral disorder in transgender persons increased rapidly during the first year of their diagnosis and, as compared to age-matched same and opposite sex controls, remain elevated throughout the study's follow-up period relative to two years prior to diagnosis. (p. 342). Correspondingly, the proportion of transgender individuals prescribed psychopharmacological agents increased significantly from baseline diagnosis. (p. 342). Notably, the odds ratio for mental and behavioral disorders increased after the index date and then gradually decreased or plateaued for transgender individuals when compared to the control population.

(p. 342). As the authors observe, "[t]his finding supports that systematic psychological evaluation after referral to a center of gender identity will increase awareness of mental health disorders in transgender individuals." (p. 342). Finally, the authors found that odds ratios were stable for mental health outcomes after initiation of gender-affirming hormone treatment. (p. 342).

12. Similar to Glintborg et al. (2023), Kaltiala (2023) used data from the general population of Finland as a control to compare the mental health outcomes for individuals with gender dysphoria. Kaltiala found that, whether or not they received medical intervention, all patients with gender dysphoria engaged with mental health services more than the cisgender controls.

13. In general, Glintborg and Kaltiala are well-done studies with reputable results, and rates of co-occurring diagnoses in these studies are aligned with previous published literature. What they show, however, is that when a person with a mental health condition initiates treatment for it, that increases the number of mental health-related visits they have. It also shows that a person who receives a mental health diagnosis will have continued treatment for that condition over time. That individual may also have other mental health diagnoses that are identified through the process of assessment and for which they will also receive treatment. Those facts do not show lack of efficacy of gender transition treatments. They simply show that when a person begins treatment for a condition, continued treatment is likely.

14. Additionally, the results must be contextualized within the broader understanding of mental health risk factors and diagnostic tends. Gender-affirming treatment does not eliminate or reduce the social stigma or bias that a person faces. We should not anticipate that outcomes for transgender individuals, even those receiving affirming treatment, will have the same baseline prevalence as the general population used for control purposes in the studies. It is also crucial to recognize that people seeking gender-affirming care are required to have full psych evaluations, thus identifying more underlying diagnoses as a part of the assessment process. If the general population were similarly subjected to a full psych evaluation, it too likely would result in a greater number of diagnoses. The best interpretation of the data, in my opinion, is that they represent the natural trajectory for diagnosis upon first contact with a full mental health evaluation. The authors agree, as discussed in the quoted language above.

15. By ignoring these critical contexts and the authors' own analysis of their data, Cantor misstates or misunderstands the clinical implications of Glintborg and Kaltiala. Cantor interprets both to support the proposition that medicalized transition is not followed by improvement in mental health. As discussed, above, improvements are not absent simply because treatment and diagnoses continue; the goal is diminishment of symptoms and improved quality of life which we see in clinical practice and which is not rebutted by this study.

16. These findings are not surprising to those of us in the medical community who interact with patients with socially stigmatized identities. Dr. Cantor's assessment that individuals that received medically affirming interventions are suffering because they continue to receive mental health services is the opposite of what I find in practice. Continued receipt of mental health services does not indicate poor mental health. Indeed, I want my patients to have access to mental health care that alleviates suffering and improves functioning. In fact, in my experience, patients who do not commit to continued care fare worse than those Cantor concludes—from his review of population-level data—are "worse off" or unwell. Treating continued care as a negative ignores the reality that mental health often requires a long term, sustained investment of time and patient participation.

17. Cantor also inappropriately interprets the studies as applicable to adolescents and children. There is little support for extrapolating to this population. While Glintborg includes adolescents and children in the sample, the authors do not analyze the data for that subset of the population separately. What the authors actually do is exclude youth from their analysis to determine if their inclusion impacts results; it does not, but this may be because the results for adults are so well powered. The exercise tells us nothing about the youth population in isolation.

18. The sample in Kaltiala is adult. Cantor discusses the difference between childhood-onset gender dysphoria and adolescent onset gender dysphoria, but he has

no first-hand knowledge of the difference. Rapid Onset Gender Dysphoria (ROGD) is a term Cantor has broadly assigned to a large group of patients he has neither seen nor assessed, and so possesses no understanding of each individual patient's time course for treatment. ROGD is not established or recognized in peer-reviewed medical literature.

3. Thompson

19. Thompson is a descriptive paper – a systematic review aimed at characterizing how care is provided. Key takeaways from the study are that (1) nothing in the study or Thompson's interpretation indicates the mental health of patients receiving gender-affirming care worsened because of care; (2) these types of studies provide only a point-in-time reference; and (3) more studies are needed to provide a more comprehensive understanding of this population and how to provide the most effective care, as is true for many other conditions. Thus, Cantor overstates the implications of the article.

20. Ultimately, Dr. Thompson's findings and recommendations are that more studies need to be conducted—not that treatment should be banned.

21. Notably, Cantor criticizes the quality of evidence regarding gender-affirming care in adolescents based on mental and physical health outcomes because "no randomized controlled trials (RCTs) yet exist" and that Thompson "included no indication that RCTs could not be conducted." (Cantor, p. 9). As with his comments

on Morandini and social transition, Cantor does not explain how he would ethically or practically conduct an RCT on medically transitioning adolescents. The selection process itself would never pass an institutional review board nor ethical muster, as patients desperately in need of and seeking care would be denied care by design. Furthermore, such a study would be impossible to blind as the physical changes associated with transitioning would be impossible to hide or disguise. Accordingly, Cantor's criticism that the evidence for providing medical treatment for transitioning adolescents is lacking because no RCTs have been conducted further demonstrates his bias and lack of experience and understanding of actual practice in the treatment of those receiving treatment for gender dysphoria.

4. Christiansen

22. Christiansen identifies factors in reducing suicidality in transgender and gender diverse youth. Transgender and gender diverse youth have higher rates of self-harm and suicide. Christiansen conducted a systematic review of 17 studies studying intervention domains: (1) crisis intervention; (2) safety and connectedness; (3) gender-affirming medical care; (4) online media; (5) family systems. As all of these studies have found, there is a developing body of literature on the impact of gender-affirming care, and many of the studies reviewed showed significant positive effects on mental and physical health. Christiansen's review points to a critical need for safety, connectedness, and acceptance from multiple systems. Cantor does not

address how targeted legislation such as the Vulnerable Child Compassion Act would meet those needs. Neither does he address the fact that Christiansen's findings indicate such legislation would further stigmatize and alienate transgender and gender diverse youth, thereby increasing the risk of suicide and self-harm. Instead, Cantor again suggests RCTs would provide high-quality studies to prove medical transition reduces suicidality. Again, Cantor's suggestion of an RCT on suicide prevention in transitioning youth raises serious questions of ethical and practical execution of such a study. By design, such a study would require withholding care from suicidal youth. As with the other RCTs suggested by Cantor, such a study would be impossible to blind. As such, Cantor's criticism of the evidence regarding the impact on the risk of suicide and self-harm in transgender youth is without merit.

B. "Lack of Science supporting the medicalized transition of minors"

23. In another attempt to argue there is no robust evidence supporting medical care for transitioning youths, Cantor turns away from peer-reviewed literature and to Twitter posts and Wall Street Journal opinion pieces. I do not consider those tweets and opinion pieces reliable sources that I would look to or rely on in forming my expert opinion.

24. Cantor also points to statements from the World Health Organization ("WHO") and the American Academy of Pediatrics ("AAP") as further "evidence" that his assessments are correct. Cantor misstates the WHO's reasoning for only

including adults by stating that minors were excluded "specifically because of the insufficient and inconsistent evidence." (Cantor, p. 28). The WHO guidance actually continues the pattern of the other studies and indicates further studies focused on long-term outcomes will assist in refining treatment guidelines and clinical practice. Nothing in the WHO's statement indicates banning gender-affirming care will provide the evidence base needed to include adolescents in the guideline.

25. Cantor's reliance on the guidance from the UK Council for Psychotherapy also is misplaced as it states the obvious to those of us actually practicing in this field: patients seeking medical interventions for gender dysphoria must be apprised of the risks associated with this treatment. Informed consent is a tenet of my practice and an ethical obligation of all medical providers. It is best practice that our patients understand to the best of their abilities at their cognitive level the risks, benefits, and potential issues associated with any course of treatment. As part of the informed consent process, I have individualized conversations with each patient—and their legal guardians, if they are younger than 18—regarding multiple aspects of treatment. Ultimately, I would like my patients to understand what they are taking, why they are taking it, and what can happen if they take it, and effort is made in practice to achieve that with each patient.

26. Cantor next criticizes the American Academy of Pediatrics' reaffirming its 2018 policy statement, despite Cantor's apparent criticism. As noted by Cantor,

the AAP has not responded to his criticisms now or in the past. It is not my place to provide an opinion as to why the AAP has not responded to Cantor's critiques, but the lack of a response from an organization such as the AAP does not strike me as evidence that it supports Cantor's position.

27. I have seen nothing in the literature I have reviewed or that Cantor reviewed that suggests banning care completely would improve outcomes. Indeed, my day-to-day experience of providing care to transitioning youth shows the opposite. In response to one of the recent bans, a patient said to me, "Why do they hate me so much?" Other patients of mine that had been in states where transition-related medical care was banned expressed fear, disappointment and a deep since of despondency that the care they have found great benefit from has been singled out to be banned.

C. WPATH SOC 8 and Appendix A

28. Cantor claims several internal communications reveal that WPATH Guideline Development Group members are concerned that young patients are being treated and harmed by "sloppy," "inexperienced," and "sometimes dangerous" providers. First, internal communications are not science and should not form the "evidence-based" opinions of a supposed expert. Second, there always is a range of quality-of-care delivery; that is why guidelines are established in order to help set the standard of care.

29. In addition, Cantor's statement that "there was no room for dissent" is the opposite of what these emails show. In my opinion, these discussions show providers engaged in a robust, transparent, and responsible discussion of how to provide the most informed and effective care patients with gender dysphoria, including consideration of divergent viewpoints, disagreements with respect to specific issues, and concerns. This is a normal, and critical, part of the scientific process and the development of practice guidelines.

II. CONCLUSION

30. I have seen nothing in the literature I have reviewed or in the State's experts' commentary that suggests banning care would improve outcomes. Instead, nearly all of the studies and commentary I have reviewed indicates banning treatment is the opposite of what is needed in this field. Additional studies, including long-term effects and impacts of medical treatment for gender dysphoria, would provide the robust evidence the State's experts claim is so desperately lacking, while banning treatment would prevent any of those desired additional studies from occurring.

31. While these studies are ongoing, rigorous debate regarding standards of care will continue. Robust debate and advocacy are essential to ensuring our patients receive the highest standards of care. I further note that, despite the States' experts' apparent dislike of it, advocacy by medical professionals on behalf of patients is not

a bad thing. Indeed, as psychiatrists, we are ethically required to advocate for our patients, and we must do the most advocacy for our distressed patients. It is self-evident that precise and careful discussions about how language of the SOC will be used in the future were not malfeasance, but prescience. To argue that such advocacy and careful deliberations about the use of language are somehow nefarious demonstrates a lack of understanding of how clinical practice actually works and of clinicians' responsibilities to their patients.

Executed this 25th day of March, 2024.

Aron Janssen, M.D.

the choice of aspirin or heparin for venous thromboembolism prophylaxis among patients with operatively treated extremity fractures (or any pelvic or acetabular fracture), this is by far the largest trial to date and provides compelling evidence that a readily available, inexpensive drug, taken orally, is a viable alternative to an injectable pharmacologic prophylaxis.

Are there any caveats to this message? The trial shows several secondary outcomes that support the main conclusion of the trial, including a similar risk of pulmonary embolism in the two groups and, in terms of safety outcomes, no evidence of a difference in the incidence of bleeding events, which occurred in 13.72% of patients in the aspirin group and 14.27% in the low-molecular-weight–heparin group. However, in keeping with previous trials, the authors noted that deep-vein thrombosis was more frequent in patients who had received aspirin than in those who had received heparin (2.51% vs. 1.71%), although the absolute difference was small (0.80 percentage points). Although deep-vein thrombosis is clearly not as serious as a fatal pulmonary embolism, it is not an inconsequential problem. Post-thrombotic syndrome affects some people who have had a deep-vein thrombosis of the leg, and this condition can cause chronic pain and swelling.[1]

The findings in this trial clearly indicate that guidelines for the prevention of hospital-acquired venous thromboembolism will need to be rewritten to include the option of aspirin in patients with traumatic injuries. More work is needed to determine whether aspirin should also be considered for venous thromboembolism prophylaxis after other types of surgeries and for nonsurgical patients who have risk factors for venous thromboembolism.

Disclosure forms provided by the author are available with the full text of this editorial at NEJM.org.

From Oxford Trauma and Emergency Care, Nuffield Department of Orthopedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom.

1. National Institute for Health and Care Excellence. Venous thromboembolism in over 16s: reducing the risk of hospital-acquired deep vein thrombosis or pulmonary embolism. August 13, 2019 (https://www.nice.org.uk/guidance/ng89/chapter/Recommendations).
2. Hegsted D, Gritsiouk Y, Schlesinger P, Gardiner S, Gubler KD. Utility of the risk assessment profile for risk stratification of venous thrombotic events for trauma patients. Am J Surg 2013;205:517-20.
3. Geerts WH, Code KI, Jay RM, Chen E, Szalai JP. A prospective study of venous thromboembolism after major trauma. N Engl J Med 1994;331:1601-6.
4. Barrera LM, Perel P, Ker K, Ciurcchi R, Farinella E, Morales Uribe CH. Thromboprophylaxis for trauma patients. Cochrane Database Syst Rev 2013;3:CD008303.
5. Rogers FB, Cipolle MD, Velmahos G, Rozycki G, Luchette FA. Practice management guidelines for the prevention of venous thromboembolism in trauma patients: the EAST Practice Management Guidelines Work Group. J Trauma 2002;53:142-64.
6. Colwell CW Jr, Pulido P, Hardwick ME, Morris BA. Patient compliance with outpatient prophylaxis: an observational study. Orthopedics 2005;28:143-7.
7. Horner D, Goodacre S, Pandor A, et al. Thromboprophylaxis in lower limb immobilisation after injury (TiLLI). Emerg Med J 2020;37:36-41.
8. Major Extremity Trauma Research Consortium (METRC). Aspirin or low-molecular-weight heparin for thromboprophylaxis after a fracture. N Engl J Med 2023;388:203-13.
9. Makedonov I, Kuhn SR, Abdulrehman J, et al. Prevention of the postthrombotic syndrome with anticoagulation: a narrative review. Thromb Haemost 2022;122:1255-64.

DOI: 10.1056/NEJMe2214045

Growing Evidence and Remaining Questions in Adolescent Transgender Care

Annelou L.C. de Vries, M.D., Ph.D., and Sabine E. Hannema, M.D., Ph.D.

This week in the *Journal*, a much-awaited primary report from Chen et al.[1] on 2 years of gender-affirming hormones (GAH) in transgender adolescents appears. The approach to adolescent transgender care with early treatment with puberty blockers, and GAH in youth from 16 years of age, originated in the Netherlands ("the Dutch model") and became the dominant medical care model for transgender adolescents.[2] Especially over the past decade, marked increases in referrals but limited evidence as to long-term outcomes have led to controversies and debate regarding this approach. Indeed, some European countries are adapting their guidelines and re-

stricting access to care for transgender youth, and some states in the United States have introduced laws to ban such care.[3] Therefore, rigorous longitudinal outcome studies that provide evidence about whether this approach is effective and safe are needed.

The results of the current study — involving a large, multisite sample of 315 participants — provide such evidence. During 24 months of GAH treatment, participant-reported appearance congruence (alignment between gender identity and physical appearance), positive affect, and life satisfaction increased and depression and anxiety decreased. In addition, initial levels and rates of change in appearance congruence correlated with the psychosocial outcomes. These results corroborate the positive effects in several earlier studies of smaller samples of adolescents and add to the evidence base that GAH can have a positive effect on mental health.[4]

Yet the study leaves some concerns unanswered. Although overall psychological functioning in the study participants improved, there was substantial variation among participants: a considerable number still had depression, anxiety, or both at 24 months, and two died by suicide. The correlation between appearance congruence and various psychological-outcome variables suggests an important mediating role of GAH and consequent bodily changes. However, other possible determinants of outcomes were not reported, particularly the extent of mental health care provided throughout GAH treatment. To date, international guidelines for transgender adolescent care recommend a psychosocial assessment and involvement of mental health professionals in a multidisciplinary care model.[5] Whether participating centers in the current study followed that approach is unfortunately unclear. Future studies that compare outcomes with different care models are needed, preferably using similar measures.

In addition, some are concerned that young persons may not be capable of making decisions regarding medical treatments that have irreversible effects that they might regret later in life. In the 2-year study by Chen et al., 9 of 314 adolescents (2.9%) stopped GAH, but it is unclear whether they detransitioned or regretted their treatment or whether they stopped because they were satisfied with treatment-related changes. Despite concerns about detransitioning, few studies have provided data on the incidence of detransitioning, and available results are inconsistent. Although one U.S. study showed that 74% of adolescents who started GAH treatment were still receiving it 4 years later, 98% of 720 Dutch adolescents who began such therapy were receiving it after a median of 2.7 years (range, 0.0 to 20.0).[6,7] Similar studies in other centers, regions, and countries are necessary to learn whether the incidence of detransitioning differs between settings and what factors are associated with these differences. It will be especially important to evaluate outcomes in adolescents starting GAH before 16 years of age, the age limit in the initial Dutch protocol.[2]

Furthermore, although Chen et al. investigated relevant psychological and gender outcome measures (e.g., depression, appearance congruence, and life satisfaction), additional factors such as autism spectrum disorder and the quality of peer relations and family support are also of interest. Social support has been hypothesized as explaining why Dutch transgender adolescents have better psychological function than those in other countries.[8] Understanding additional factors that influence outcomes should help to determine which components of care and support other than GAH might improve the lives of transgender adolescents.

Finally, benefits of early medical intervention, including puberty suppression, need to be weighed against possible adverse effects — for example, with regard to bone and brain development and fertility. At present, studies involving young adults from the Dutch adolescent transgender cohort show that accrual of bone mineral decelerates during puberty suppression but increases during GAH treatment and also that adolescents' educational achievements are as expected given their pretreatment status, which is reassuring.[9,10] However, those results from a single Dutch center should be replicated and validated in other contexts, as in a sample followed in the current study.

Despite uncertainties that call for further study, current information shows that mental health improves with GAH, whereas withholding treatment may lead to increased gender dysphoria and adversely affect psychological functioning. The study by Chen et al. adds to the

evidence of the effectiveness of the current care model that includes hormonal treatment for transgender adolescents.

Disclosure forms provided by the authors are available with the full text of this editorial at NEJM.org.

From the Departments of Child and Adolescent Psychiatry (A.L.C.V.) and Pediatrics (S.E.H.), Center of Expertise on Gender Dysphoria, Amsterdam University Medical Centers, Location Vrije Universiteit, Amsterdam.

1. Chen D, Berona J, Chan Y-M, et al. Psychosocial functioning in transgender youth after 2 years of hormones. N Engl J Med 2023;388:240-50.
2. Delemarre-van de Waal HA, Cohen-Kettenis PT. Clinical management of gender identity disorder in adolescents: a protocol on psychological and paediatric endocrinology aspects. Eur J Endocrinol 2006;155:Suppl 1:S131-S137.
3. Cass Review. Independent review of gender identity services for children and young people: interim report. February 2022 (https://cass.independent-review.uk/publications/interim-report/).
4. Kuper LE, Stewart S, Preston S, Lau M, Lopez X. Body dissatisfaction and mental health outcomes of youth on gender-affirming hormone therapy. Pediatrics 2020;145(4):e20193006.
5. Coleman E, Radix AE, Bouman WP, et al. Standards of care for the health of transgender and gender diverse people, version 8. Int J Transgend Health 2022;23:Suppl 1:S1-S259.
6. Roberts CM, Klein DA, Adirim TA, Schvey NA, Hisle-Gorman E. Continuation of gender-affirming hormones among transgender adolescents and adults. J Clin Endocrinol Metab 2022;107(9):e3937-e3943.
7. van der Loos MATC, Hannema SE, Klink DT, den Heijer M, Wiepjes CM. Continuation of gender-affirming hormones in transgender people starting puberty suppression in adolescence: a cohort study in the Netherlands. Lancet Child Adolesc Health 2022;6:869-75.
8. de Graaf NM, Steensma TD, Carmichael P, et al. Suicidality in clinic-referred transgender adolescents. Eur Child Adolesc Psychiatry 2022;31:67-83.
9. Schagen SEE, Schagen SEE, Wouters FM, Cohen-Kettenis PT, Gooren LJ, Hannema SE. Bone development in transgender adolescents treated with GnRH analogues and subsequent gender-affirming hormones. J Clin Endocrinol Metab 2020;105(12):e4252-e4263.
10. Arnoldussen M, Hooijman EC, Kreukels BP, de Vries AL. Association between pre-treatment IQ and educational achievement after gender-affirming treatment including puberty suppression in transgender adolescents. Clin Child Psychol Psychiatry 2022;27:1069-76.

DOI: 10.1056/NEJMe2216191