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 3
2. Executive summary of the review 4
Critical outcomes 4
Important outcomes 6
Important outcomes 7
Discussion 13
Conclusion 14
3. Methodology 14
Review questions 14
Review process 15
4. Summary of included studies 15
5. Results 21
6. Discussion 47
7. Conclusion 50
Appendix A PICO 51
Appendix B Search strategy 55
Appendix C Evidence selection 70
Appendix D Excluded studies table 70
Appendix E Evidence tables 77
Appendix F Quality appraisal checklists 107
Appendix G Grade profiles 109
Glossary 153
References 155

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