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[3] The CBCL/YSR has 2 components: Internalising score which sums the anxious/depressed, withdrawn-depressed, and somatic complaints scores; externalising score which sums rule-breaking and aggressive behaviour.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

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: 0.157 (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

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