Puberty blockers for youth experiencing gender dysphoria: A systematic review and meta-analysis

Corresponding author: Anna Miroshnychenko, MSc, PhD(c)
HSC-2C, Department of Health Research Methods, Evidence and Impact, McMaster University
1280 Main Street West, Hamilton, ON, Canada, L8S 4L8
Email: mirosha@mcmaster.ca

Word Count: Abstract: 221; Total: 2999
Number of Tables: 3
Number of Figures: 1
Number of Appendices: 14

Puberty blockers for youth experiencing gender dysphoria: A systematic review and meta-analysis

Authors: Anna Miroshnychenko, MSc; Yetiani M Roldan, MD; Sara Ibrahim, BHSc; Chan Kulatunga-Moruzi, MSc, PhD; Steven Montante, MD; Rachel Couban, PhD; Gordon Guyatt, MD, MSc; Romina Brignardello-Petersen, DDS, MSc, PhD

Abstract:

Aim. Gender dysphoria (GD) refers to the psychological distress associated with the incongruence between one's sex and one's gender identity. To manage gender dysphoria (GD), individuals may delay the development of primary and secondary sex characteristics with the use of puberty blockers. In this systematic review, we assess and summarize the certainty of the evidence about the effects of puberty blockers in individuals experiencing GD.

Methods: We searched MEDLINE, Embase, PsychINFO, Social Sciences Abstracts, LGBTQ+ Source, and Sociological Abstracts through September 2023. We included observational studies comparing puberty blockers to no puberty blockers in individuals under 26 years of age experiencing GD, as well as before-after and case series studies. Outcomes of interest included psychological and physical. Pairs of reviewers independently screened articles, abstracted data, and assessed risk of bias. We performed meta-analysis and assessed the certainty of a non-zero effect using the GRADE approach.

Results: We included 10 studies. Comparative observational studies (n=3), comparing puberty blockers versus no puberty blockers, provided very low certainty evidence on the outcomes of global function and depression. Before-after studies (n=7) provided very low certainty evidence addressing gender dysphoria, global function, depression, and bone mineral density.

Conclusion: There remains considerable uncertainty regarding the effects of puberty blockers in individuals experiencing GD. Methodologically rigorous prospective studies are needed to elucidate the effects of this intervention.

Key messages:

1. *What is known on this topic*: Previously published systematic reviews addressing the effects of puberty blockers in individuals experiencing GD have not conducted a meta-analysis.
2. *What this study adds*: This publication addresses the effects of puberty blockers in individuals experiencing GD, while adhering to the highest methodological standards for conducting and reporting a systematic review and meta-analysis and assessing the risk of bias in each included study and the certainty of the evidence for each outcome of interest.
3. *How this study might affect research, practice, and policy*: The evidence from this systematic review and meta-analysis can be used to inform individuals experiencing GD and considering puberty blockers, clinicians involved in their care as well as clinical practice guideline developers, policy makers and stakeholders who make decisions about treatment related to gender dysphoria.

Key words: puberty blockers, puberty suppression, gender dysphoria

Introduction

Gender dysphoria refers to intense psychological distress or impairment in functioning attributed to the feelings of incongruence between one's gender identity and sex assigned at birth [1]. Individuals experiencing gender dysphoria may seek hormonal and surgical interventions to align their bodies with their experienced or expressed gender. These interventions, including hormonal treatments or surgeries, aim to alleviate the distress caused by GD and improve mental well-being. [2]

Puberty blockers, or gonadotropin-releasing hormone analogues (GnRHa), suppress the release of sex hormones an delay puberty's physical changes, which normally begins between the ages of 8 and 13 for natal females and between the ages of 9 and 14 for natal males and follows a 5-stage process. [3] Initially developed to treat precocious puberty, these medications have more recently been used to manage gender dysphoria. [4,5] By pausing puberty, it was postulated that they would provide time for individuals to explore their gender identity without the added stress of unwanted secondary sexual characteristics, before deciding whether to continue with gender-affirming hormone therapy.[6,7] While originally considered fully reversible, [7-9] concerns have emerged about potential long-term effects and partial irreversibility. [10,11]

The use of puberty blockers in gender dysphoria remains controversial due to the methodological limitations of previously published evidence syntheses and individual studies. [12-14] This systematic review, using the highest methodological standards, synthetizes the evidence to inform decision-making regarding puberty blockers for youth with gender dysphoria.

Methods

We report this systematic review and meta-analysis following the guidance of the preferred reporting items for systematic reviews and meta-analyses (PRISMA) checklist (Appendix 1). We registered the protocol in PROSPERO (registration ID: CRD42023452171).

Eligibility criteria

For the eligibility criteria, see Appendix 2.

Information sources

With the assistance of an information specialist (RC), we searched in MEDLINE, Embase, PsycINFO, Social Sciences Abstracts, Contemporary Women's Issues, LGBTQ+ Source, Sociological Abstracts, Studies on Women, Gender Abstracts, and Google Scholar from inception to September 2023. The search for this systematic review was part of an umbrella search for another related systematic review. [15] All search strategies are included in Appendix 3.

Study selection

Using Covidence software (https://www.covidence.org/), a pair of reviewers (SI, YR), following training and calibration exercises, independently screened titles and abstracts, and full texts of potentially eligible studies. A third reviewer (AM) resolved conflicts. The study selection was completed in tandem alongside another related systematic review at the abstract and full text stages. [15]

Data collection

For data collection, see Appendix 4.

Risk of bias in included studies

For each eligible study and outcome, a pair of reviewers (SI, YR), following training and calibration exercises, used a modified version of the Cochrane risk of bias tool for non-randomized studies of interventions (ROBINS-I) [16] to ensure standardized and consistent assessments across study designs (i.e., studies comparing two groups, studies comparing before-after, and case series). Reviewer rated studies as having low, moderate, high, or critical risk of bias across several domains (Appendix 5; Appendix 6). For randomized control trials (RCTs), we planned to use the revised Cochrane risk of bias tool. [17] Reviewers resolved discrepancies by discussion or by consulting a third reviewer (AM) when necessary.

Data synthesis

While study authors used various observational study designs, we classified studies as comparative observational if they reported outcome data for an intervention group compared to an independent group. We considered studies as before-after if researchers measured outcomes in a single group before and after the intervention, and as case series if researchers measured outcomes in a single group after the intervention. Depending on how outcomes were measured and reported, studies could be classified under different designs for different outcomes.

For dichotomous outcomes, we summarized the effect of interventions using odds ratios in comparative observational and before-after studies, and proportions (i.e., number of events per number of participants in the study group) in case series. For continuous outcomes, we summarized the effects of interventions using mean difference in comparative observational studies (i.e., difference in scores between the study groups), mean change in before-after studies (i.e., difference in scores before and after intervention), and mean in case series.

Since the study authors did not provide correlation coefficients, we imputed a moderate correlation coefficient (r=0.5) when calculating mean change. We calculated 95% confidence intervals (CI) around all estimates.

We conducted meta-analysis using a random-effects model when appropriate, according to subject area experts (CKM, SM), of studies addressing the same outcome and if there was no clinical heterogeneity between them (i.e., study design, population, intervention/comparator, outcome definition). When two or more studies reported the same outcome using different scales, we reported the effect estimate as a standardized mean change for before-after studies.

When we could not perform a meta-analysis, we provided summaries of evidence across studies for each outcome. We used the *meta* and *metafor* packages in R Studio Version 4.2 for analyses.

Certainty of the evidence

We assessed the certainty of the evidence using the grading of recommendations assessment, development, and evaluation (GRADE) approach. [18] For each comparison and outcome, a pair of methodologists with experience in GRADE (SI, YR) rated each domain independently, resolving discrepancies by consulting a third methodologist (AM). We rated the certainty as high, moderate, low, or very low. All bodies of evidence started as high certainty, [19] and could be rated down for risk of bias, inconsistency, indirectness, imprecision, and publication bias; or rated up when a large magnitude of effect or a dose-response relationship was observed, or when all plausible confounders or other biases increased our confidence in the estimated effect. [20]

Following GRADE guidance, when assessing risk of bias at the outcome level, we rated down the certainty of the evidence up to three levels for risk of prognostic imbalance in observational comparative studies where risk of bias at the study level was assessed using the ROBINS-I tool. [19] For case series, we rated down three levels due to lack of a comparison group.

To minimize value judgments, we used a null effect threshold (1 for relative measures, and 0 for absolute measures and mean differences or mean changes) to rate the certainty that puberty blockers caused any benefit or harm, regardless of magnitude. We did not establish a minimally important difference to infer whether an effect was important or not. We assessed the causal

effect of puberty blockers on health outcomes, rather than associations, even if the included studies were not design with this aim. Following GRADE guidance and principles to address questions about interventions using observational studies, we defined the target question, [21] clarified its intent (causality), and assessed the certainty of the evidence.[22] We used GRADEpro to create the summary of findings tables.[23]

Subgroup and sensitivity analyses

For the subgroup and sensitivity analyses, see Appendix 7 and 8.

Management of conflicts of interest

For the management of conflicts of interest, see Appendix 9. Other systematic reviews that are part of the described agreement include systematic reviews about the effects of social gender transition (submitted for publication), mastectomy,[24] chest binding and genital tucking (submitted for publication), and gender-affirming hormone therapy (submitted for publication).

Results

After screening 6,736 titles and abstracts for this systematic review and another related systematic review, [15] we included 10 studies in this systematic review. Figure 1 shows the results of the study search and selection process. We present the reasons for exclusion (n=311) with references in Appendix 10.

Characteristics of included studies

Of 10 included studies, 3 were comparative observational and 7 used a before-after design (Figure 1). [8,25-33] In addition, two of the before-after studies reported data about progression to gender-affirming hormone therapy after the intervention and we classified these as case series for that outcome. [27,30] After conducting the search, we did not identify any RCTs meeting our eligibility criteria.

The mean (SD) age of participants at the time of puberty blockers ranged from 12.93 (2.52) to 16.48 (1.26). We present characteristics of included studies in Appendix 11. Appendix 12 describes measurement instruments and their interpretability.

Risk of bias in included studies

Across comparative observational studies, the domain most frequently judged as serious or critical risk of bias were confounding and missing data. Before-after studies were at serious or critical risk of bias due to missing data, and moderate or critical risk of bias due to deviation from intended intervention and lack of an independent comparator group. Case series were at critical risk of bias due to deviation from intended intervention (i.e., administration of co-interventions) and lack of a comparison group (Appendix 6).

Effects of puberty blockers

We describe the effects of the intervention for each study design (i.e., comparative observational studies, before-after study design, case series). Tables 1-3 provide summary of findings tables. Appendix 13 displays forest plots of meta-analysis.

1. Comparative observational studies

Global function: When assessed at 12 months with the Children's Global Assessment Scale ranging from 1 to 100 (higher scores = greater global function), the meta-analysis suggests that the difference in mean change in scores from baseline (MC) may be higher (MC 7.67 higher [95% CI 2 lower to 17.34 higher], number of studies (n) = 2, very low certainty) in individuals who received puberty blockers compared to those who do not, although we are very uncertain about the causal effect of the intervention on global function. When assessed at 6 months, the evidence about global function was also very low certainty. See Table 1.

Depression: When measured at 12 months with the Center for Epidemiologic Studies Depression Scale (CESD-R) ranging from 0 to 60 (higher scores = greater depression), a linear regression analysis reported that puberty blockers may not decrease depression scores in female to male participants (r^2=0.09, b=-0.02, p=0.95), but may decrease depression in male to female (r^2=0.52, b=-2.41, p=0.008) participants. We are very uncertain about the causal effect of the intervention on depression. See Table 1.

Table 1. Puberty blockers vs no puberty blockers: evidence from comparative observational studies.

Outcomes	Anticipated absolute effects* (95% CI)		№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	No puberty blockers	**Puberty blockers**			
Global Function, Long Term Follow-Up assessed with: participant reported Children's Global Assessment Scale Scale from: 1 to 100, higher scores = greater global function follow-up: 12 months [a]	-	difference in mean change from baseline **7.67 higher** (2 lower to 17.34 higher)	103 (2 non-randomised studies)[1,2]	⨁◯◯◯ Very low [b,c,d]	The evidence is very uncertain about the effect of puberty blockers on global function, at long term follow-up.
Global Function, Short Term Follow-Up assessed with: participant reported Children's Global Assessment Scale Scale from: 1 to 100, higher scores = greater global function follow-up: 6 months [e]	-	difference in mean change from baseline **0.36 lower** (0.96 lower to 0.24 higher)	121 (1 non-randomised study)[2]	⨁◯◯◯ Very low [f,g]	The evidence is very uncertain about the effect of puberty blockers on global function at short term follow-up.
Depression, Long term Follow-Up assessed with: participant reported The Center for Epidemiologic Studies Depression Scale; Scale from: 1 to 60, higher scores = greater depression follow-up: 12 months [a]	88% participants received puberty blockers. A linear regression analysis reported that, when measuring depression using the CESDS-R, and using as the reference no puberty blockers, puberty blockers: - did not result in a statistically significant decrease in scores in female to male participants (R2= 0.09, b= -0.02, p= 0.95); - resulted in a statistically significant decrease in score in male to female participants (R2 =0.52, b= -2.41, p= 0.008). The analysis adjusted for psychiatric medications and engagement in counseling.		26 (1 non-randomised study)[3]	⨁◯◯◯ Very low [h,i]	The evidence is very uncertain about the effect of puberty blockers on depression at long term follow-up.
Other outcomes - not measured [j]	-	-	-	-	-

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: confidence interval

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Explanations

a. Long Term Follow-Up: outcome measured at ≥ 12 months follow-up.
b. Rated down three levels for critical risk of bias due to lack of adjustment for important confounders (psychiatric interventions, mental health comorbidities, socioeconomic status, or family support) and missing data (less than 50% provided outcome data) in 2 included studies.
c. Statistically, there was considerable heterogeneity with I2=99% and p<0.01. However, we did not rate down for inconsistency as the overall effect estimate was not importantly affected by the studies contributing to statistical heterogeneity.
d. Rated down one level for imprecision due to the confidence interval (CI) crossing the threshold of no effect (i.e., difference in mean change from baseline=0) and a CI that suggests both a possibility of a benefit or a harm in the outcome.
e. Short Term Follow-Up: outcome measured at ≤ 6 months follow-up.
f. Rated down three levels for critical risk of bias due to lack of adjustment for important confounders (psychiatric interventions, mental health comorbidities, socioeconomic status, or family support) and serious risk of bias due to missing data (i.e., 60% of participants provided outcome data).
g. Optimal information size (OIS) of 200 participants was not met as only 121 participants were included in this study. Rated down one levels for imprecision because of this. Low sample size importantly increases the risk of random error.
h. Rated down two levels for serious risk of bias due to lack of adjustment for important confounders (i.e., mental health comorbidities, socioeconomic status, or family support) and missing data (i.e., 56% of participants provided outcome data).
i. Optimal information size (OIS) of 200 participants was not met as only 26 participants were included in this study. Rated down two levels for imprecision because of this. Low sample size importantly increases the risk of random error.
j. Outcomes not measured: gender dysphoria, death by suicide, sexual dysfunction, progression to gender-affirming hormone treatment, bone mineral density.

References

1.I, Becker-Hebly, S, Fahrenkrug, F, Campion, H, Richter-Appelt, M, Schulte-Markwort, C, Barkmann. Psychosocial health in adolescents and young adults with gender dysphoria before. European child & adolescent psychiatry; 2021.
2.R, Costa, M, Dunsford, E, Skagerberg, V, Holt, P, Carmichael, M, Colizzi. Psychological Support, Puberty Suppression, and Psychosocial Functioning in. The journal of sexual medicine; 2015.
3.C, Achille, T, Taggart, NR, Eaton, J, Osipoff, K, Tafuri, A, Lane, TA, Wilson. Longitudinal impact of gender-affirming endocrine intervention on the mental. International journal of pediatric endocrinology; 2020.

2. Before-after studies

Gender dysphoria: When measured between 23 to 36 months with the Utrecht Gender Dysphoria Scale ranging from 1 to 5 (higher scores = greater gender dysphoria), meta-analysis suggests that gender dysphoria may be lower (SMC (standardized mean change) 0.01 lower [95% CI 0.4 lower to 0.19 higher], n = 2, very low certainty) after receiving puberty blockers compared to before, although we are very uncertain about the causal effect of the intervention on gender dysphoria. See Table 2.

Global function: When measured between 23 to 36 months with the Children's Clinical Global Assessment ranging from 1 to 100 (higher scores = greater global function), meta-analysis suggests that global function may be higher (MC 3.63 higher [95% CI 3.17 higher to 4.09

higher], n = 2, very low certainty) after receiving puberty blockers compared to before, although we are very uncertain about the causal effect of the intervention on global function. See Table 2.

Depression: When measured at 23 months with the Beck Depression Inventory ranging from 0 to 63 (higher scores = greater depression), depression may be lower (MC 3.36 lower [95% CI 3.69 lower to 3.03 lower], n = 1, very low certainty) after receiving puberty blockers compared to before. See Table 2.

Bone mineral density of the hip: When assessed between 12 and 36 months with dual energy x-ray absorptiometry (DXA), z-scores ranging from -3 to 3, meta-analysis suggests that bone density of the hip may be lower (MC 0.71 lower [95% CI 1.09 lower to 0.33 lower], n = 2, very low certainty) after receiving puberty blockers compared to before, although we are very uncertain about the causal effect of the intervention on bone mineral density. See Table 2.

Bone mineral density of the lumbar spine: When assessed between 12 and 36 months with dual energy x-ray absorptiometry (DXA), z-scores ranging from -3 to 3, meta-analysis suggests that bone density of the lumbar spine may be lower (MC 0.72 lower [95% CI 0.91 lower to 0.54 lower], n = 5, very low certainty) after receiving puberty blockers compared to before, although we are very uncertain about the causal effect of the intervention on bone mineral density. When assessed at 6 months, the evidence about this outcome was also very low certainty. See Table 2.

Bone mineral density of the femoral neck: When assessed between 20 and 24 months with dual energy x-ray absorptiometry (DXA), z-scores ranging from -3 to 3, meta-analysis suggests

that bone density of the femoral neck may be lower (MC 0.7 lower [95% CI 1.11 lower to 0.29 lower], n = 2, very low certainty) after receiving puberty blockers compared to before, although we are very uncertain about the causal effect of the intervention on bone mineral density. See Table 2.

Table 2. Puberty blockers vs no puberty blockers: evidence from before-after studies.

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	No puberty blockers	**Puberty blockers**				
Gender Dysphoria, Long Term Follow-Up assessed with: participant reported Utrecht Gender Dysphoria Scale Scale from: 1 to 5, higher scores = greater gender dysphoria follow-up: range 23 months to 36 months [a]	-	standardized mean change **0.1 lower** (0.4 lower to 0.19 higher)	-	59 (2 non-randomised studies)[1,2]	⨁◯◯◯ Very low [b,c,d]	The evidence is very uncertain about the effect of puberty blockers on gender dysphoria at long term follow-up.
Global Function, Long Term Follow-Up assessed with: participant reported Children's Clinical Global Assessment Scale from: 1 to 100, higher scores = greater global function follow-up: range 23 months to 36 months [a]	The mean global function, long term follow-up was **66.53**	mean change **3.63 higher** (3.17 higher to 4.09 higher)	-	53 (2 non-randomised studies)[1,2]	⨁◯◯◯ Very low [b,e]	The evidence is very uncertain about the effect of puberty blockers on global function at long term follow-up.
Depression, Long Term Follow-Up assessed with: participant reported Beck Depression Inventory Scale from: 0 to 63, higher scores = greater depression follow-up: 23 months [a]	The mean depression, long term follow-up was **8.31**	mean change **3.36 lower** (3.69 lower to 3.03 lower)	-	41 (1 non-randomised study)[1]	⨁◯◯◯ Very low [e,f]	The evidence is very uncertain about the effect of puberty blockers on depression at long term follow-up.
Bone Mineral Density - Hip, Long Term Follow-Up assessed with: Dual-energy x-ray absorptiometry (DXA), z-scores Scale from: -3 to 3 follow-up: range 12 months to 36 months [a]	The mean bone mineral density - hip, long term follow-up was **-0.02**	mean change **0.71 lower** (1.09 lower to 0.33 lower)	-	128 (2 non-randomised studies)[3,4]	⨁◯◯◯ Very low [g,h]	The evidence is very uncertain about the effect of puberty blockers on bone mineral density - hip at long term follow-up.

Bone Mineral Density - Lumbar Spine, Long Term Follow-Up assessed with: Dual-energy x-ray absorptiometry (DXA), z-scores Scale from: -3 to 3 follow-up: range 12 months to 36 months [a]	The mean bone mineral density - lumbar spine, long term follow-up was -**0.13**	mean change **0.72 lower** (0.91 lower to 0.54 lower)	-	222 (5 non-randomised studies)[3,4,5,6,7]	⨁◯◯◯ Very low [i,j]	The evidence is very uncertain about the effect of puberty blockers on bone mineral density - lumbar spine at long term follow-up.
Bone Mineral Density - Lumbar Spine, Short Term Follow-Up assessed with: Dual-energy x-ray absorptiometry (DXA), z-scores Scale from: -3 to 3 follow-up: 6 months [k]	The mean bone mineral density - lumbar spine, short term follow-up was -**1**	mean change **1.3 lower** (1.57 lower to 1.03 lower)	-	9 (1 non-randomised study)[6]	⨁◯◯◯ Very low [l,m]	The evidence is very uncertain about the effect of puberty blockers on bone mineral density - lumbar spine at short term follow-up.
Bone Mineral Density - Femoral Neck, Long Term Follow-up assessed with: Dual-energy x-ray absorptiometry (DXA), z-scores Scale from: -3 to 3 follow-up: range 20 months to 24 months [a]	The mean bone mineral density - femoral neck, long term follow-up was -**0.43**	mean change **0.7 lower** (1.11 lower to 0.29 lower)	-	93 (2 non-randomised studies)[5,7]	⨁◯◯◯ Very low [e,n,o]	The evidence is very uncertain about the effect of puberty blockers on bone mineral density - femoral neck at long term follow-up.
Other outcomes - not measured [p]	-	-	-	-	-	

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: confidence interval

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Explanations

a. Long Term Follow-Up: outcome measured at ≥ 12 months follow-up.
b. Rated down three levels for risk of bias due to risk of bias with respect to prognostic imbalance associated with the observational study design not having a comparison group, and one of the two included studies having critical risk of bias because of deviation from intended intervention (i.e., all participants received social support and psychotherapy) and missing data (i.e., 43.9% participants provided outcome data).
c. Rated down one level for inconsistency due to heterogeneity among included studies marked by a lack of overlap of confidence intervals between the included studies. Statistically, there is considerable heterogeneity with I2=96% and p<0.01.
d. Rated down two levels for imprecision as the confidence intervals cross the threshold of no effect (i.e., difference in mean change from baseline=0), suggesting both a possibility of a benefit or a harm in the outcome; and because the optimal information size (OIS) of 200 participants was not met (i.e., low sample size importantly increases the risk of random error).
e. Rated down one level for imprecision because the optimal information size of 200 participants was not met (<100 participants included). Low sample size importantly increases the risk of random error.
f. Rated down two levels due to risk of bias stemming from prognostic imbalance associated with the observational study design not having a comparison group.
g. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design not having a comparison group. Moreover, of the two included studies, one has critical risk and the second serious risk due to missing outcome data (i.e., 9.92% and 68.24%, respectively, provided outcome data).
h. Statistically, there was considerable heterogeneity with I2=97% and p<0.01. However, we did not rate down for inconsistency as the overall effect estimate was not importantly affected by the studies contributing to statistical heterogeneity.
i. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design not having a comparison group. Moreover, of the five included studies, three have critical risk of bias (i.e., 28.7%, 10.74%, 27.27%, respectively, provided outcome data), one has serious risk (i.e. 68.24% provided outcome data), and another one has moderate risk (i.e., 85.9% provided outcome data) due to missing outcome data.
j. Statistically, there was considerable heterogeneity with I2=89% and p<0.01. However, we did not rate down for inconsistency as the overall effect estimate was not importantly affected by the studies contributing to statistical heterogeneity.
k. Short Term Follow-Up: outcome measured at ≤ 6 months follow-up.
l. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design not having a comparison group. Moreover, the study has critical risk due to missing outcome data (i.e., 27.27% provided outcome data).
m. Rated down two levels for imprecision because the optimal information size of 200 participants was not met (6 participants included). Low sample size importantly increases the risk of random error.
n. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design not having a comparison group. Moreover, one of two included studies has critical risk (i.e., 44.29% provided outcome data) and another one has serious risk (i.e., 60% provided outcome data) of bias due to missing outcome data.
o. Statistically, there was considerable heterogeneity with I2=98% and p<0.01. However, we did not rate down for inconsistency as the overall effect estimate was not importantly affected by the studies contributing to statistical heterogeneity.
p. Outcomes not measured: death by suicide, sexual dysfunction from a psychological perspective (i.e., lack of erection, dyspareunia, problems related to dry and degenerated mucosal tissue, anorgasmia), progression to gender-affirming hormone treatment.

References

1.AL, de,Vries, TD, Steensma, TA, Doreleijers, PT, Cohen-Kettenis. Puberty suppression in adolescents with gender identity disorder: prospective. The journal of sexual medicine; 2011.
2.P, Carmichael, G, Butler, U, Masic, TJ, Cole, BL, De,Stavola, S, Davidson, EM, Skageberg, S, Khadr, RM, Viner. Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year. PloS one; 2021.
3.SEE, Schagen, FM, Wouters, PT, Cohen-Kettenis, LJ, Gooren, SE, Hannema. Bone Development in Transgender Adolescents Treated With GnRH Analogues and. The Journal of clinical endocrinology and metabolism; 2020.
4.B, Navabi, K, Tang, K, Khatchadourian, ML, Lawson. Pubertal Suppression, Bone Mass, and Body Composition in Youth With Gender. Pediatrics; 2021.
5.D, Klink, M, Caris, A, Heijboer, M, van,Trotsenburg, J, Rotteveel. Bone mass in young adulthood following gonadotropin-releasing hormone analog. The Journal of clinical endocrinology and metabolism; 2015.
6.E, Karakılıç,Özturan, AP, Öztürk, F, Baş, AB, Erdoğdu, S, Kaptan, AD, Kardelen,Al, Ş, Poyrazoğlu, M, Yıldız, N, Direk, Ş, Yüksel, F, Darendeliler. Endocrinological Approach to Adolescents with Gender Dysphoria: Experience of a. Journal of clinical research in pediatric endocrinology; 2023.
7.T, Joseph, J, Ting, G, Butler. The effect of GnRH analogue treatment on bone mineral density in young. Journal of pediatric endocrinology & metabolism: JPEM; 2019.

3. Case series

Two of the before-after studies reported data about progression to gender-affirming hormone therapy after the intervention and we classified these as case series for that outcome [27,30].

Progression to gender-affirming hormone therapy: Within a range of 12 to 36 months, 92% of individuals who received puberty blockers progressed to receiving gender-affirming hormone therapy (proportion 0.92 [95% CI 0.53 to 0.99], n = 2, very low certainty), although we are very uncertain about the effects of the intervention on this outcome. When assessed at 12 months, the evidence about this outcome was also very low certainty. See Table 3. In terms of the incidence of this outcome after receiving puberty blockers, the certainty of the evidence is low. See Appendix 14.

Table 3. Puberty blockers vs no puberty blockers: evidence from case series. *

Outcomes	Anticipated absolute effects# (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with no puberty blockers	**Risk with puberty blockers**				
Progression to Gender-Affirming Hormone Therapy, Long Term Follow-Up assessed with: data from medical records follow-up: range 12 months to 36 months [a]	No comparison group available	**920 per 1,000** (530 to 990)	**proportion 0.92** (0.53 to 0.99)	65 (2 non-randomised studies)[1,2]	⨁◯◯◯ Very low [b,c,d]	The evidence is very uncertain about the effect of puberty blockers on progression to gender affirming hormone therapy at long term follow-up.
Progression to Gender-Affirming Hormone Therapy, Short Term Follow-Up assessed with: data from medical records follow-up: 12 months [a]	No comparison group available	**690 per 1,000** (390 to 910)	**proportion 0.69** (0.39 to 0.91)	13 (1 non-randomised study)[1]	⨁◯◯◯ Very low [b,d]	The evidence is very uncertain about the effect of puberty blockers on progression to gender affirming hormone therapy at long term follow-up.
Other outcomes - not measured [e]	-	-	-	-	-	

#**The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: confidence interval

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Explanations

Outcomes	Anticipated absolute effects# (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with no puberty blockers	**Risk with puberty blockers**				
a. Long Term Follow-Up: outcome measured at ≥ 12 months follow-up. b. Rated down three levels due to lack of a comparison group when assessing the effect of puberty blockers on progression of gender-affirming hormone therapy. We did not rate down for risk of bias due to deviation from intended intervention (i.e., all participants were receiving psychosocial support and psychiatric interventions), because these co-interventions would likely result in less individuals receiving the intervention of interest. c. Statistically, there was considerable heterogeneity with I2=74% and p<0.01. However, we did not rate down for inconsistency as the overall effect estimate was not importantly affected by the studies contributing to statistical heterogeneity. d. Rated down two levels for imprecision because the optimal information size (OIS) of 200 participants was not met. Low sample size importantly increases the risk of random error. e. Outcomes not measured: gender dysphoria, death by suicide, global function, depression, sexual dysfunction from physiological perspective (i.e., lack of erection, dyspareunia, problems related to dry and degenerated mucosal tissue, anorgasmia), bone mineral density.						
References						
1.E, Karakılıç,Özturan, AP, Öztürk, F, Baş, AB, Erdoğdu, S, Kaptan, AD, Kardelen,Al, Ş, Poyrazoğlu, M, Yıldız, N, Direk, Ş, Yüksel, F, Darendeliler. Endocrinological Approach to Adolescents with Gender Dysphoria: Experience of a. Journal of clinical research in pediatric endocrinology; 2023. 2.P, Carmichael, G, Butler, U, Masic, TJ, Cole, BL, De,Stavola, S, Davidson, EM, Skageberg, S, Khadr, RM, Viner. Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year. PloS one; 2021.						

*The research question of interest involves the comparison of puberty blockers to no puberty blockers.

Discussion

This systematic review and meta-analysis synthesized and appraised the available evidence regarding the effects of puberty blockers in youth experiencing gender dysphoria. Most studies provided very low certainty evidence about the outcomes of interest, thus, we cannot exclude the possibility of benefit or harm.

Although some may consider our modification of the ROBINS-I tool for assessing risk of bias a limitation, we believe that this adjustment produced conclusions comparable to those that would have been reached using the original tool or alternative tools, such as the Newcastle-Ottawa scale. [34] Methodological shortcomings in the included studies would likely yield similar findings across any risk of bias tool. Comparative observational studies had critical risk bias due to confounding and missing data. Before-after studies had moderate to critical risk of bias due to missing data and moderate to critical risk of bias due to deviation from intended intervention. In addition to lacking a comparison group, case series studies were at critical risk of bias due to

deviation from intended intervention (i.e., administration of co-interventions). Given their design, findings from case series studies should only be used for hypothesis generation.

To address the target question of this systematic review and that of the decision makers of whether these interventions should be used, we evaluated the effects of puberty blockers using case series and before-after studies because randomized clinical trials and comparative observational studies were unavailable. While these study designs can provide insights for certain single-group questions (e.g., what is the quality of life of individuals who have received puberty blockers), they cannot answer questions about the effects of interventions (e.g., whether quality of life is better in individuals who received puberty blockers compared to those who did not). It is crucial to account for these limitations when the target question focuses on intervention effects. Therefore, we rated down the certainty of the evidence primarily because of risk of bias and imprecision for most outcomes and study designs. Imprecision often resulted from an insufficient sample size and confidence intervals crossing the null effect threshold. We did not find data for the outcomes of death by suicide and sexual dysfunction.

This is the first systematic review and meta-analysis to assess the effects of puberty blockers in children, adolescents, and young adults experiencing gender dysphoria using the highest methodological standards. [35] Several other published systematic reviews have assessed puberty blockers, and their conclusions align with ours. [36-40] One of these systematic reviews used the ROBINS-I tool, [36] while others used a different tool to assess the risk of bias. [37-40] Only two of these systematic reviews assessed the certainty of the evidence using GRADE guidance,[36,37] and none conducted a meta-analysis. All other published systematic reviews had similar conclusions

to our review: the current best available evidence about the effects of puberty blockers in the population of interest is very low certainty, and high-quality studies evaluating short and long-term outcomes of puberty blockers are needed.

To understand the effects of puberty blockers in youth with gender dysphoria, methodologically rigorous studies, such as RCTs (if ethical) and prospective cohort studies, are needed to produce higher certainty evidence. Since the current best evidence, including our systematic review and meta-analysis is predominantly very low certainty, clinicians must clearly communicate this evidence to patients and caregivers. Treatment decisions should consider the lack of moderate- and high-quality evidence, uncertainty about the effects of puberty blockers, and patient's values and preferences Given the individualistic nature of values and preferences, guideline developers and policy makers should be transparent about which and whose values they are prioritizing when making recommendations and policy decisions.

Strengths and limitations of the review process

This systematic review and meta-analysis has multiple strengths. We rigorously followed the highest methodological standards, assessed the risk of bias for each study, and evaluated the certainty of the evidence for each outcome using the latest guidance. We performed analyses and interpreted results following the GRADE approach. A limitation of our review is the inclusion of only English language studies. However, we do not expect this to fundamentally change our conclusions. Additionally, due to feasibility considerations, we had to prioritize outcomes to

include in our systematic review. Therefore, we cannot make any conclusions regarding other outcomes of interest, such as regret, anxiety, and pelvic pain.

Conclusion

The best available evidence reporting the effects of puberty blockers in youth with gender dysphoria was mostly very low certainty and therefore we cannot exclude the possibility of benefit or harm. There was evidence available for the outcomes of global function, depression, gender dysphoria, bone mineral density, and progression to gender-affirming hormone therapy. High certainty evidence from prospective cohort studies and, if ethical, RCTs, is needed to understand the short- and long-term effects of puberty blockers in individuals experiencing gender dysphoria.

Contributorship Statement:

AM contributed to the conception and design, data collection, analysis and interpretation, and drafted and critically revised the manuscript. YR and SI contributed to data collection, analysis and interpretation, and critically revised the manuscript. CKM contributed to the conception and design and critically revised the manuscript. SM contributed to the conception and design, data interpretation, and critically revised the manuscript. RC contributed to data collection. GG critically revised the manuscript. RBP contributed to conception and design, data interpretation, and critically revised the manuscript. RBP is the guarantor of this work.

Funding and Disclosures:

The funding and disclosures statement includes all authors of this manuscript as well as all authors of the published protocol. The authors of the published protocol include representatives from the sponsor who participated only in the development of the systematic review question.

1. Funding:

This work was commissioned by the Society for Evidence-based Gender Medicine (SEGM), the sponsor, and McMaster University. This systematic review is part of a large research project funded through a research agreement between the Society for Evidence-based Gender Medicine (SEGM), the sponsor, and McMaster University. None of the team members received financial compensation directly from SEGM to conduct this systematic review and meta-analysis.

2. Financial conflicts of interest (manuscript authors):

Direct financial conflicts of interest:

Dr. Romina Brignardello-Petersen and Anna Miroshnychenko provided methodological expertise for the SEGM initiative to summarize and appraise the quality of publications related to gender medicine for the SEGM online platform, and for this work they received financial compensation from SEGM. This work was completely independent from the systematic review and meta-analysis.

Indirect financial conflicts of interest:

Manuscript authors do not have indirect financial conflicts of interest.

2. Financial conflicts of interest (as reported by the protocol authors who were not part of the evidence synthesis team at the time of their participation in the generation of the question):

Direct financial conflicts of interest:

E. Abbruzzese is a contributing author for the Society for Evidence-based Gender Medicine online platform and received financial compensation from SEGM.

Indirect financial conflicts of interest:

Dr. William Malone's fee for publishing a research article as “open access” was compensated by SEGM.

3. Other disclosures (manuscript authors):

Expressed opinions:

Dr. Chan Kulatunga-Moruzi has expressed opinions on recommendations for gender-affirming care for transgender and gender-diverse youth in the Journal of Pediatrics and Child Health. This opinion piece was published after this systematic review was submitted for publication, and the content of this systematic review did not change.

3. Other disclosures (as reported by the protocol authors who were not part of the evidence synthesis team at the time of their participation in the generation of the question):

Affiliations:

Dr. William Malone is a board member of SEGM.

Expressed opinions:

Dr. William Malone has expressed opinions about gender affirmation interventions for adolescents and young adults in The Journal of Clinical Endocrinology and Metabolism, The Lancet, Child and Adolescent Health, and Medscape.

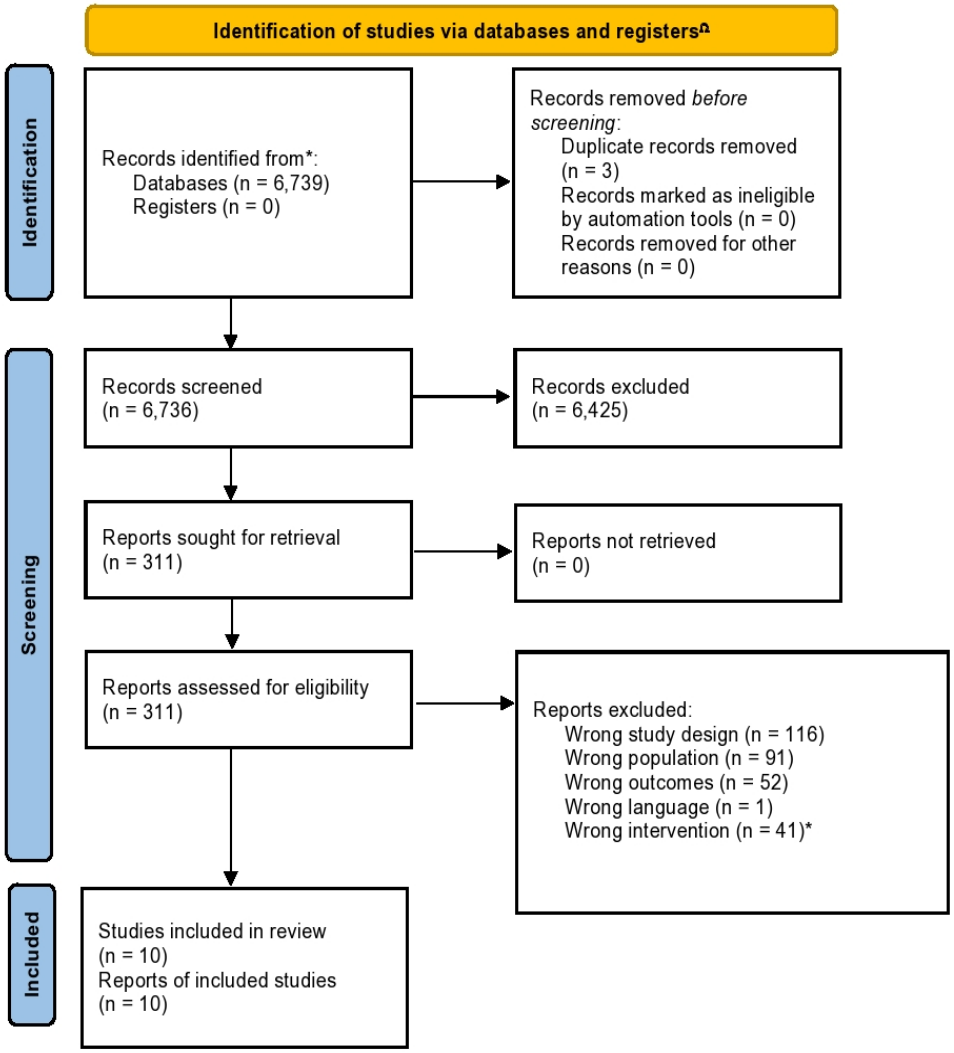
References

1. American Psychiatric Association. The Diagnostic and Statistical Manual of Mental Disorders (5th ed DSM-5). Arlington, VA: American Psychiatric Association; 2013.
2. Trans Care BC. 2024. https://www.transcarebc.ca/.
3. Breehl L, Caban O. Physiology, Puberty. StatPearls. Treasure Island (FL) ineligible companies. Disclosure: Omar Caban declares no relevant financial relationships with ineligible companies.: StatPearls Publishing
Copyright © 2024, StatPearls Publishing LLC.; 2024.
4. Bangalore Krishna K, Fuqua JS, Rogol AD, et al. Use of Gonadotropin-Releasing Hormone Analogs in Children: Update by an International Consortium. *Horm Res Paediatr* 2019; **91**(6): 357-72.
5. Carel JC, Eugster EA, Rogol A, et al. Consensus statement on the use of gonadotropin-releasing hormone analogs in children. *Pediatrics* 2009; **123**(4): e752-62.
6. Delemarre-van de Waal HA, Cohen-Kettenis PT. Clinical management of gender identity disorder in adolescents: a protocol on psychological and paediatric endocrinology aspectsThis paper was presented at the 4th Ferring Pharmaceuticals International Paediatric Endocrinology Symposium, Paris (2006). Ferring Pharmaceuticals has supported the publication of these proceedings. *European Journal of Endocrinology* 2006; **155**(Supplement_1): S131-S7.
7. Gooren L, Delemarre-van de Waal H. The Feasibility of Endocrine Interventions in Juvenile Transsexuals. *Journal of Psychology & Human Sexuality* 1996; **8**(4): 69-74.
8. de Vries AL, Steensma TD, Doreleijers TA, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. *J Sex Med* 2011; **8**(8): 2276-83.
9. NICE NIfHaCE. Evidence review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria. 2020.
10. Baxendale S. The impact of suppressing puberty on neuropsychological function: A review. *Acta Paediatr* 2024; **113**(6): 1156-67.
11. Cass H. Independent review of gender identity swrvices for chidlren and young people: Final report. 2024: 29.
12. Coleman E, Radix AE, Bouman WP, et al. Standards of Care for the Health of Transgender and Gender Diverse People, Version 8. *Int J Transgend Health* 2022; **23**(Suppl 1): S1-s259.
13. Hembree WC, Cohen-Kettenis PT, Gooren L, et al. Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society* Clinical Practice Guideline. *The Journal of Clinical Endocrinology & Metabolism* 2017; **102**(11): 3869-903.
14. Rew L, Young CC, Monge M, Bogucka R. Review: Puberty blockers for transgender and gender diverse youth-a critical review of the literature. *Child Adolesc Ment Health* 2021; **26**(1): 3-14.
15. Miroshnychenko A, Ibrahim S, Roldan Y, et al. Gender-affirming hormone therapy for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis *Submitted for publication* 2024.
16. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016; **355**: i4919.
17. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019; **366**: l4898.

18. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; **336**(7650): 924-6.
19. Schünemann HJ, Cuello C, Akl EA, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *J Clin Epidemiol* 2019; **111**: 105-14.
20. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 2011; **64**(12): 1311-6.
21. Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings. *Journal of Clinical Epidemiology* 2021; **137**: 163-75.
22. Brignardello-Petersen R, Guyatt GH. Assessing the Certainty of the Evidence in Systematic Reviews: Importance, Process, and Use. *Am J Epidemiol* 2024.
23. GRADEpro GDT. GRADEpro Guideline Development Tool. McMaster University and Evidence Prime; 2022.
24. Miroshnychenko A, Roldan YM, Ibrahim S, et al. "Mastectomy for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis". *Plast Reconstr Surg* 2024.
25. Achille C, Taggart T, Eaton NR, et al. Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: preliminary results. *Int J Pediatr Endocrinol* 2020; **2020**: 8.
26. Becker-Hebly I, Fahrenkrug S, Campion F, Richter-Appelt H, Schulte-Markwort M, Barkmann C. Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: a descriptive study from the Hamburg Gender Identity Service. *Eur Child Adolesc Psychiatry* 2021; **30**(11): 1755-67.
27. Carmichael P, Butler G, Masic U, et al. Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK. *PLoS One* 2021; **16**(2): e0243894.
28. Costa R, Dunsford M, Skagerberg E, Holt V, Carmichael P, Colizzi M. Psychological Support, Puberty Suppression, and Psychosocial Functioning in Adolescents with Gender Dysphoria. *J Sex Med* 2015; **12**(11): 2206-14.
29. Joseph T, Ting J, Butler G. The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort. *J Pediatr Endocrinol Metab* 2019; **32**(10): 1077-81.
30. Karakılıç Özturan E, Öztürk AP, Baş F, et al. Endocrinological Approach to Adolescents with Gender Dysphoria: Experience of a Pediatric Endocrinology Department in a Tertiary Center in Turkey. *J Clin Res Pediatr Endocrinol* 2023; **15**(3): 276-84.
31. Klink D, Caris M, Heijboer A, van Trotsenburg M, Rotteveel J. Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. *J Clin Endocrinol Metab* 2015; **100**(2): E270-5.
32. Navabi B, Tang K, Khatchadourian K, Lawson ML. Pubertal Suppression, Bone Mass, and Body Composition in Youth With Gender Dysphoria. *Pediatrics* 2021; **148**(4).
33. Schagen SEE, Wouters FM, Cohen-Kettenis PT, Gooren LJ, Hannema SE. Bone Development in Transgender Adolescents Treated With GnRH Analogues and Subsequent Gender-Affirming Hormones. *J Clin Endocrinol Metab* 2020; **105**(12): e4252-63.
34. Wells GA, Wells, G., Shea, B., Shea, B., O'Connell, D., Peterson, J., Welch, Losos, M., Tugwell, P., Ga, S.W., Zello, G.A., & Petersen, J.A. . The Newcastle-Ottawa Scale (NOS) for

assessing the quality of nonrandomized studies in meta-analyses. Ottawa Hospital Research Institute. 2014.
35. Higgins J, Thomas J, (editors). Cochrane Handbook for Systematic Reviews of Interventions Cochrane; 2023.
36. Ludvigsson JF, Adolfsson J, Höistad M, Rydelius PA, Kriström B, Landén M. A systematic review of hormone treatment for children with gender dysphoria and recommendations for research. *Acta Paediatr* 2023; **112**(11): 2279-92.
37. National Institute for Health and Care Excellence. Evidence review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria. 2020.
38. Taylor J, Mitchell A, Hall R, et al. Interventions to suppress puberty in adolescents experiencing gender dysphoria or incongruence: a systematic review. *Archives of Disease in Childhood* 2024: archdischild-2023-326669.
39. Thompson L, Sarovic D, Wilson P, et al. A PRISMA systematic review of adolescent gender dysphoria literature: 3) treatment. *PLOS Glob Public Health* 2023; **3**(8): e0001478.
40. Zepf FD, König L, Kaiser A, et al. [Beyond NICE: Updated Systematic Review on the Current Evidence of Using Puberty Blocking Pharmacological Agents and Cross-Sex-Hormones in Minors with Gender Dysphoria]. *Z Kinder Jugendpsychiatr Psychother* 2024; **52**(3): 167-87.

Figure 1. PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only



[Ω] This was an umbrella search completed for two related systematic reviews and meta-analyses. Ten studies were included in this systematic review. The studies that were included in another review are part of the studies excluded for wrong intervention.

*Twenty-four of 41 studies excluded for wrong intervention were included in another review.

Source: Page MJ, et al. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.