



OPEN ACCESS

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

Anna Miroshnychenko ¹ , ¹ Yetiani Roldan, ¹ Sara Ibrahim, ¹ Chan Kulatunga-Moruzi, ² Steven Montante, ³ Rachel Couban, ⁴ Gordon Guyatt, ⁵ Romina Brignardello-Petersen ⁵

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/archdischild-2024-327909>).

¹Health Research Methods, Evidence, and Impact, McMaster University Faculty of Health Sciences, Hamilton, Ontario, Canada

²Independent Researcher, Hamilton, Ontario, Canada

³Montante Plastic Surgery and Aesthetics, Richmond, Virginia, USA

⁴McMaster University Faculty of Health Sciences, Hamilton, Ontario, Canada

⁵Department of Health Research Methods, Evidence, and Impact, McMaster University Faculty of Health Sciences, Hamilton, Ontario, Canada

Correspondence to

Dr Gordon Guyatt;
guyatt@mcmaster.ca

Received 28 August 2024

Accepted 4 January 2025

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 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 summarise 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 from inception to September 2023. We included observational studies comparing puberty blockers with no puberty blockers in individuals aged <26 years experiencing GD, as well as before–after and case series studies. Outcomes of interest included psychological and physical outcomes. Pairs of reviewers independently screened articles, abstracted data and assessed risk of bias. We performed a meta-analysis and assessed the certainty of a non-zero effect using the grading of recommendations assessment, development and evaluation (GRADE) approach.

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

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

Trial registration number PROSPERO
CRD42023452171.

INTRODUCTION

Gender dysphoria (GD) 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 GD 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 wellbeing.²

Puberty blockers, or gonadotropin releasing hormone analogues, suppress the release of sex hormones and delay puberty's physical changes, which normally begins between the ages of 8 and

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Previously published systematic reviews addressing the effects of puberty blockers in individuals with gender dysphoria (GD) have not conducted a meta-analysis.

WHAT THIS STUDY ADDS

⇒ This study addressed the effects of puberty blockers in individuals with GD, while adhering to the highest methodological standards for conducting and reporting a systematic review and meta-analysis.
⇒ The risk of bias in each included study and the certainty of the evidence for each outcome of interest were assessed.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE AND POLICY

⇒ The evidence from this systematic review and meta-analysis can be used to inform individuals with 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 GD.

13 years for natal females and between the ages of 9 and 14 for natal males, and follows a five 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 the 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} In this systematic review, using the highest methodological standards, we synthesised the evidence to inform decision making regarding puberty blockers for individuals with gender dysphoria.

METHODS

We report this systematic review and meta-analysis following the guidance of the Preferred Reporting



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

To cite: Miroshnychenko A, Roldan Y, Ibrahim S, et al. *Arch Dis Child* Epub ahead of print: [please include Day Month Year]. doi:10.1136/archdischild-2024-327909

Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (online supplemental appendix 1).

Eligibility criteria

For eligibility criteria, see online supplemental appendix 2.

Information sources

With the assistance of an information specialist (RC), we searched 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 online supplemental appendix 3.

Study selection

Using Covidence software (<https://www.covidence.org/>), a pair of reviewers (SI, YR), after 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 alongside another related systematic review at the abstract and full text stages.¹⁵

Data collection

For data collection, see online supplemental appendix 4.

Risk of bias in included studies

For each eligible study and outcome, a pair of reviewers (SI, YR), after training and calibration exercises, used a modified version of the Cochrane risk of bias tool for non-randomised studies of interventions (ROBINS-I)¹⁶ to ensure standardised and consistent assessments across study designs (ie, 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 (online supplemental appendices 5 and 6). For randomised 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 the authors of the studies used various observational study designs, we classified studies as comparative observational if they reported outcome data for an intervention group compared with 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 summarised the effect of interventions using ORs in comparative observational and before–after studies and proportions (ie, number of events per number of participants in the study group) in case series. For continuous outcomes, we summarised the effects of interventions using mean difference in comparative observational studies (ie, difference in scores between the study groups), mean change in before–after studies (ie, difference in scores before and after the intervention) and mean in case series. Because the authors of the studies did not provide correlation coefficients, we imputed a moderate correlation coefficient ($r=0.5$) when calculating mean change. We calculated 95% CI around all estimates.

We conducted a meta-analysis using a random effects model when appropriate, according to subject area experts (CK-M, SM), of studies addressing the same outcome and if there was no clinical heterogeneity between them (ie, study design, population, intervention/comparator or outcome definition). When two or more studies reported the same outcome using different scales, we reported the effect estimate as a standardised 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 V.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. Evidence could also be 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 minimise 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 designed 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 subgroup and sensitivity analyses, see online supplemental appendices 7 and 8.

Management of conflicts of interest

For the management of conflicts of interest, see online supplemental appendix 9. Other systematic reviews that are part of the described agreement included 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 6736 titles and abstracts for this systematic review and another related systematic review,¹⁵ we included 10 studies in our review. Figure 1 shows the results of the study search and

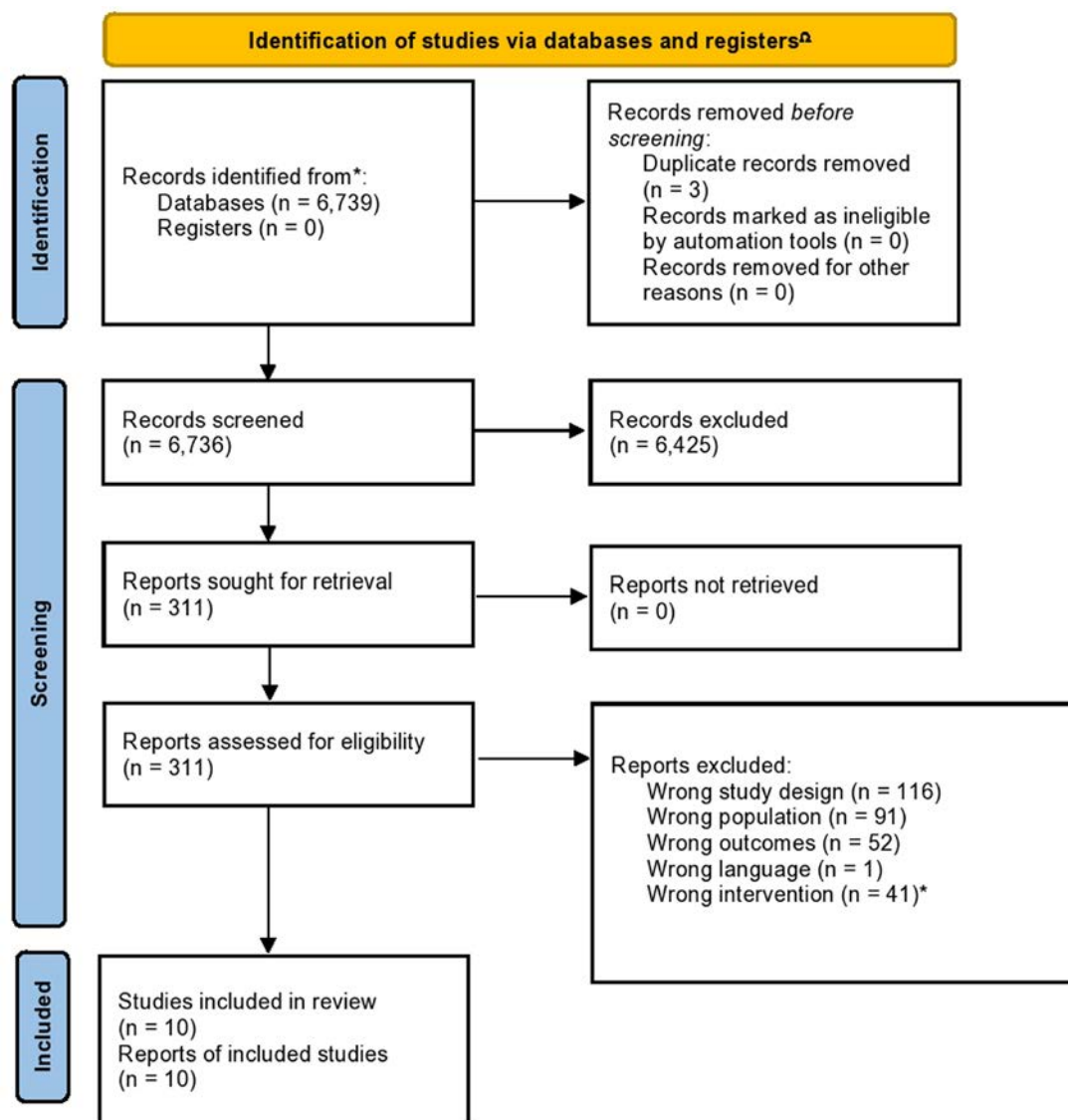


Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow diagram for new systematic reviews that included searches of databases and registers only. ^QThis was an umbrella search completed for two related systematic reviews and meta-analyses. Ten studies were included 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 *et al.*⁴⁰

selection process. We present the reasons for exclusion (n=311) with references in online supplemental appendix 10.

Characteristics of included studies

Of the 10 included studies, three were comparative observational and seven 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.

Mean age of participants at the time of puberty blockers ranged from 12.93 (SD 2.52) to 16.48 (1.26) years. The characteristics of the included studies are presented in online supplemental appendix 11. Online supplemental appendix 12 describes the measurement instruments and their interpretability.

Risk of bias in included studies

Across comparative observational studies, the domains 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 (ie, administration of co-interventions) and lack of a comparison group (online supplemental appendix 6).

Effects of puberty blockers

We described the effects of the intervention for each study design (ie, comparative observational studies, before–after study design and case series). Tables 1–3 provide a summary of the findings. Online supplemental appendix 13 displays forest plots of the meta-analysis.

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

Outcomes	Anticipated absolute effects* (95% CI)		No of participants (studies)	Certainty of evidence (GRADE)	Comments
	No puberty blockers	Puberty blockers			
Global function, long term follow-up assessed with: participant reported Children's Global Assessment Scale (1–100), higher scores=greater global function. Follow-up: 12 months†	–	Difference in mean change from baseline 7.67 higher (2 lower to 17.34 higher)	103 (2 non-randomised studies) ^{26 28}	⊕○○○ Very low‡§¶	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 (1–100), higher scores=greater global function. Follow-up: 6 months**	–	Difference in mean change from baseline 0.36 lower (0.96 lower to 0.24 higher)	121 (1 non-randomised study) ²⁸	⊕○○○ Very low††‡	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 Centre for Epidemiologic Studies Depression Scale (CESDS-R, 1–60), higher scores=greater depression. Follow-up: 12 months†	88% of participants received puberty blockers. A linear regression analysis reported that, when measuring depression with 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 ($R^2=0.09$, $b=-0.02$, $p=0.95$); resulted in a statistically significant decrease in score in male to female participants ($R^2=0.52$, $b=-2.41$, $p=0.008$). The analysis adjusted for psychiatric medications and engagement in counselling		26 (1 non-randomised study) ²⁵	⊕○○○ Very low§§¶¶	Evidence is very uncertain about the effect of puberty blockers on depression at long term follow-up
Other outcomes, not measured***	–	–	–	–	–

Grading of recommendations assessment, development and evaluation (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.

*Risk in the intervention group (with 95% CI) based on the assumed risk in the comparison group and the relative effect of the intervention (with 95% CI).

†Long term follow-up: outcome measured at ≥ 12 months of follow-up.

‡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 (<50% provided outcome data) in two included studies.

§Statistically, there was considerable heterogeneity with $I^2=99\%$ 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.

¶Rated down one level for imprecision due to the CI crossing the threshold of no effect (ie, difference in mean change from baseline=0) and a CI that suggests both a possibility of a benefit or a harm in the outcome.

**Short term follow-up: outcome measured at ≤ 6 months follow-up.

††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 (ie, 60% of participants provided outcome data).

‡‡Optimal information size of 200 participants was not met as only 121 participants were included in this study. Rated down one level for imprecision because of this. Low sample size importantly increases the risk of random error.

§§Rated down two levels for serious risk of bias due to lack of adjustment for important confounders (ie, mental health comorbidities, socioeconomic status or family support) and missing data (ie, 56% of participants provided outcome data).

¶¶Optimal information size 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.

***Outcomes not measured: gender dysphoria, death by suicide, sexual dysfunction, progression to gender affirming hormone treatment and bone mineral density.

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 suggested 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), $n=2$ studies, very low certainty) in individuals who received puberty blockers compared with those who did 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 (table 1).

Depression

When measured at 12 months with the Centre 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 participants ($r^2=0.52$,

$b=-2.41$, $p=0.008$). We are very uncertain about the causal effect of the intervention on depression (table 1).

Before–after studies

Gender dysphoria

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

Global function

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

Table 2 Puberty blockers versus no puberty blockers: evidence from before–after studies

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Certainty of evidence (GRADE)	Comments
	No puberty blockers	Puberty blockers				
Gender dysphoria, long term follow-up assessed with: participant reported Utrecht Gender Dysphoria Scale (1–5), higher scores=greater gender dysphoria. Follow-up: 23–36 months†	–	Standardised mean change 0.1 lower (0.4 lower to 0.19 higher)	–	59 (2 non-randomised studies) ^{8,27}	⊕○○○ Very low‡§¶	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 (1–100), higher scores=greater global function. Follow-up: 23–36 months†	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) ^{8,27}	⊕○○○ Very low‡**	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 (0–63), higher scores=greater depression. Follow-up: 23 months†	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) ⁸	⊕○○○ Very low**††	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: DXA, z scores scale (–3 to 3). Follow-up: 12–36 months†	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) ^{32,33}	⊕○○○ Very low‡§§	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: DXA, z scores scale (–3 to 3). Follow-up: 12–36 months†	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) ^{29–33}	⊕○○○ Very low¶¶***	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: DXA, z scores scale (–3 to 3). Follow-up: 6 months†††	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) ³⁰	⊕○○○ Very low‡§§§§	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: DXA, z scores scale (–3 to 3). Follow-up: 20–24 months†	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) ^{29,31}	⊕○○○ Very low**¶¶¶****	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††††	–	–	–	–	–	

Grading of recommendations assessment, development and evaluation (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 the effect.

*Risk in the intervention group (with 95% CI) based on the assumed risk in the comparison group and the relative effect of the intervention (with 95% CI).

†Long term follow-up: outcome measured at ≥12 months of follow-up.

‡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 a critical risk of bias because of deviation from intended intervention (ie, all participants received social support and psychotherapy) and missing data (ie, 43.9% participants provided outcome data).

§Rated down one level for inconsistency due to heterogeneity among included studies marked by a lack of overlap of CIs between the included studies. Statistically, there was considerable heterogeneity, with $I^2=96\%$ and $p<0.01$.

¶Rated down two levels for imprecision as the CIs crossed the threshold of no effect (ie, 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 of 200 participants was not met (ie, low sample size importantly increases the risk of random error).

***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.

††Rated down two levels due to risk of bias stemming from prognostic imbalance associated with the observational study design not having a comparison group.

‡‡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 had a critical risk and the second a serious risk due to missing outcome data (ie, 9.92% and 68.24%, respectively, provided outcome data).

§§Statistically, there was considerable heterogeneity with $I^2=97\%$ 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.

¶¶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 had a critical risk of bias (ie, 28.7%, 10.74%, 27.27%, respectively, provided outcome data), one had a serious risk (ie, 68.24% provided outcome data) and another one had a moderate risk (ie, 85.9% provided outcome data) due to missing outcome data.

****Statistically, there was considerable heterogeneity with $I^2=89\%$ 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.

†††Short term follow-up: outcome measured at ≤6 months follow-up.

‡‡‡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 had a critical risk due to missing outcome data (ie, 27.27% provided outcome data).

§§§Rated down two levels for imprecision because the optimal information size of 200 participants was not met (six participants included). Low sample size importantly increases the risk of random error.

¶¶¶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 had a critical risk (ie, 44.29% provided outcome data) and another one had a serious risk (ie, 60% provided outcome data) of bias due to missing outcome data.

****Statistically, there was considerable heterogeneity with $I^2=98\%$ 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.

††††Outcomes not measured: death by suicide, sexual dysfunction from a psychological perspective (ie, lack of erection, dyspareunia, problems related to dry and degenerated mucosal tissue, and anorgasmia) and progression to gender affirming hormone treatment.

DXA, dual energy x-ray absorptiometry.

higher to 4.09 higher), $n=2$ studies, very low certainty) after receiving puberty blockers compared with before, although we are very uncertain about the causal effect of the intervention on global function (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

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

Outcomes	Anticipated absolute effects† (95% CI)		Relative effect (95% CI)	No of participants (studies)	Certainty of 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: 12–36 months‡	No comparison group available	920 per 1000 (530 to 990)	Proportion 0.92 (0.53 to 0.99)	65 (2 non-randomised studies) ^{27 30}	⊕○○○ Very low§¶**	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‡	No comparison group available	690 per 1000 (390 to 910)	Proportion 0.69 (0.39 to 0.91)	13 (1 non-randomised study) ²⁷	⊕○○○ Very low§**	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††	–	–	–	–	–	

Grading of recommendations assessment, development and evaluation (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.

*Research question of interest involves comparison of puberty blockers with no puberty blockers.

†Risk in the intervention group (with 95% CI) based on the assumed risk in the comparison group and the relative effect of the intervention (with 95% CI).

‡Long term follow-up: outcome measured at ≥12 months follow-up.

§Outcomes not measured: gender dysphoria, death by suicide, global function, depression, sexual dysfunction from physiological perspective (ie, lack of erection, dyspareunia, problems related to dry and degenerated mucosal tissue, and anorgasmia) and bone mineral density.

¶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 (ie, all participants were receiving psychosocial support and psychiatric interventions), because these co-interventions would likely result in less individuals receiving the intervention of interest.

**Statistically, there was considerable heterogeneity with $I^2=74\%$ 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.

††Rated down two levels for imprecision because the optimal information size of 200 participants was not met. Low sample size importantly increases the risk of random error.

to 3.03 lower), $n=1$ study, very low certainty) after receiving puberty blockers compared with before (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 suggested that bone density of the hip may be lower (MC 0.71 lower (95% CI 1.09 lower to 0.33 lower), $n=2$ studies, very low certainty) after receiving puberty blockers compared with before, although we are very uncertain about the causal effect of the intervention on bone mineral density (table 2).

Bone mineral density of the lumbar spine

When assessed between 12 and 36 months with DXA, z scores ranging from -3 to 3 , meta-analysis suggested that bone density of the lumbar spine may be lower (MC 0.72 lower (95% CI 0.91 lower to 0.54 lower), $n=5$ studies, very low certainty) after receiving puberty blockers compared with 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 (table 2).

Bone mineral density of the femoral neck

When assessed between 20 and 24 months with DXA, z scores ranging from -3 to 3 , meta-analysis suggested that bone density of the femoral neck may be lower (MC 0.7 lower (95% CI 1.11 lower to 0.29 lower), $n=2$ studies, very low certainty) after receiving puberty blockers compared with before, although we are very uncertain about the causal effect of the intervention on bone mineral density (table 2).

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 12–36 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), $n=2$ studies, 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 (table 3). In terms of the incidence of this outcome after receiving puberty blockers, the certainty of the evidence was low (online supplemental appendix 14).

DISCUSSION

This systematic review and meta-analysis synthesised and appraised the available evidence regarding the effects of puberty blockers in youths with GD. Most studies provided very low certainty of evidence about the outcomes of interest, and 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 with 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 give similar findings across any risk of bias tool. Comparative observational studies had a 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 (ie, 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 randomised clinical trials and comparative observational studies were unavailable. While these study designs can provide insights for certain single group questions (eg, what is the quality of life of individuals who have received puberty blockers), they cannot answer questions about the effects of interventions (eg, whether quality of life is better in individuals who received puberty blockers compared with 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 due to 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 with GD using the highest methodological standards.³⁵ Several other published systematic reviews have assessed puberty blockers and their conclusions align with ours.^{9 36–39} One of these systematic reviews used the ROBINS-I tool,³⁶ while others used a different tool to assess the risk of bias.^{9 37–39} Only two of these systematic reviews assessed the certainty of the evidence using GRADE guidance,^{9 36} 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 individuals with GD, 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, was 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 prioritising 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 was 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 prioritise outcomes for inclusion 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 individuals with GD 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, GD, 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 GD.

Contributors 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. CK-M 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. RB-P contributed to conception and design, data interpretation and critically revised the manuscript. RB-P is the guarantor of this work.

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.

Disclaimer 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.

Competing interests Direct financial conflicts of interest: RB-P and AM provided methodological expertise for the Society for Evidence-based Gender Medicine (SEGM) initiative to summarise 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 independent of the systematic review and meta-analysis. 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): E Abbruzzese is a contributing author for the SEGM online platform and received financial compensation from SEGM; William Malone's fee for publishing a research article as 'open access' was compensated by SEGM. Other disclosures (manuscript authors): CKM 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. 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): William Malone is a board member of SEGM. William Malone has expressed opinions about gender affirmation interventions for adolescents and young adults in *Journal of Clinical Endocrinology and Metabolism*, *The Lancet*, *Child and Adolescent Health* and *Medscape*.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially,

and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Anna Miroshnychenko <http://orcid.org/0000-0003-1788-6234>

REFERENCES

- American Psychiatric Association. *The Diagnostic and Statistical Manual of Mental Disorders*. 5th edn. Arlington, VA: American Psychiatric Association, 2013.
- Trans care BC. 2024. Available: <https://www.transcarebc.ca>
- Breehl L, Caban O. Physiology, puberty. In: *StatPearls*. Treasure Island (FL) ineligible companies. Disclosure: Omar Caban declares no relevant financial relationships with ineligible companies: StatPearls Publishing LLC, 2024. Available: <https://www.ncbi.nlm.nih.gov/books/NBK534827/#article-27994.s9>
- 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;91:357–72.
- Carel J-C, Eugster EA, Rogol A, et al. Consensus statement on the use of gonadotropin-releasing hormone analogs in children. *Pediatrics* 2009;123:e752–62.
- Cohen-Kettenis PT. Clinical management of gender identity disorder in adolescents: a protocol on psychological and paediatric endocrinology aspects. This paper was presented at the 4th Ferring Pharmaceuticals International Paediatric Endocrinology Symposium. *Eur J Endocrinol* 2006;155:S131–7.
- Gooren L, Delemarre-van de Waal H. The Feasibility of Endocrine Interventions in Juvenile Transsexuals. *J Psychol Human Sex* 1996;8:69–74.
- de Vries ALC, Steensma TD, Doreleijers TAH, et al. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. *J Sex Med* 2011;8:2276–83.
- National Institute for Health and Care Excellence. Evidence review: gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria. 2020. Available: https://cass.independent-review.uk/wp-content/uploads/2022/09/20220726_Evidence-review_GnRH-analogues_For-upload_Final.pdf
- Baxendale S. The impact of suppressing puberty on neuropsychological function: A review. *Acta Paediatr* 2024;113:1156–67.
- Cass H. Independent review of gender identity services for children and young people: final report. 2024.29. Available: <https://cass.independent-review.uk/home/publications/final-report/>
- 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:S1–259.
- Hembree WC, Cohen-Kettenis PT, Gooren L, et al. Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society* Clinical Practice Guideline. *J Clin Endocrinol Metab* 2017;102:3869–903.
- Rew L, Young CC, Monge M, et al. Review: Puberty blockers for transgender and gender diverse youth—a critical review of the literature. *Child Adolesc Ment Health* 2021;26:3–14.
- 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. *Arch Dis Child* 2024.
- 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.
- Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:14898.
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- 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.
- Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 2011;64:1311–6.
- Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings. *J Clin Epidemiol* 2021;137:163–75.
- Brignardello-Petersen R, Guyatt GH. Assessing the Certainty of the Evidence in Systematic Reviews: Importance, Process, and Use. *Am J Epidemiol* 2024;:kwae332.
- GRADEpro GDT. GRADEpro guideline development tool. McMaster University and Evidence Prime; 2022. Available: <https://www.gradepr.org/>
- 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.
- 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.
- Becker-Hebly I, Fahrenkrug S, Campion F, et al. 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;30:1755–67.
- 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:e0243894.
- Costa R, Dunsford M, Skagerberg E, et al. Psychological Support, Puberty Suppression, and Psychosocial Functioning in Adolescents with Gender Dysphoria. *J Sex Med* 2015;12:2206–14.
- 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;32:1077–81.
- 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;15:276–84.
- Klink D, Caris M, Heijboer A, et al. 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:E270–5.
- Navabi B, Tang K, Khatchadourian K, et al. Pubertal Suppression, Bone Mass, and Body Composition in Youth With Gender Dysphoria. *Pediatrics* 2021;148:e2020039339.
- Schagen SEE, Wouters FM, Cohen-Kettenis PT, et al. Bone Development in Transgender Adolescents Treated With GnRH Analogues and Subsequent Gender-Affirming Hormones. *J Clin Endocrinol Metab* 2020;105:dgaa604:e4252–63.
- Wells GA, Shea BJ, O'Connell D, et al. The newcastle-ottawa scale (nos) for assessing the quality of nonrandomized studies in meta-analyses. 2014. Available: https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
- Higgins J, Thomas J, eds. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane, 2023. Available: <https://training.cochrane.org/handbook>
- Ludvigsson JF, Adolfsson J, Höistad M, et al. A systematic review of hormone treatment for children with gender dysphoria and recommendations for research. *Acta Paediatr* 2023;112:2279–92.
- Taylor J, Mitchell A, Hall R, et al. Interventions to suppress puberty in adolescents experiencing gender dysphoria or incongruence: a systematic review. *Arch Dis Child* 2024;109:s33–47.
- 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:e0001478.
- 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;52:167–87.
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.