Appendix Attachment

1a

ATTACHMENT A

RON DESANTIS
GOVERNOR

SIMONE MARSTILLER
SECRETARY

April 20, 2022

Tom Wallace
Deputy Secretary for Medicaid
Agency for Health Care Administration
2727 Mahan Drive
Tallahassee, FL 32308

Dear Deputy Secretary Wallace:

On April 20, 2022, the Florida Department of Health released guidance on the treatment of gender dysphoria for children and adolescents.[1] The Florida Medicaid program does not have a policy on whether to cover such treatments for Medicaid recipients diagnosed with gender dysphoria. Please determine, under the process described in Florida Administrative Code Rule 59G-1035, whether such treatments are consistent with generally accepted professional medical standards and not experimental or investigational. Pursuant to Rule 59G-1035(5), I look forward to receiving your final determination.

Sincerely,

Simone Marstiller
Secretary

[1] *See* https://www.floridahealth.gov/newsroom/2022/04/20220420-gender-dysphoria-press-release.pr.html (last visited Apr. 20, 2022).

Appendix Attachment

1b

ATTACHMENT B

59G-1.035 Determining Generally Accepted Professional Medical Standards.

(1) Definitions.

(a) Generally accepted professional medical standards – Standards based on reliable scientific evidence published in peer-reviewed scientific literature generally recognized by the relevant medical community or practitioner specialty associations' recommendations.

(b) Health service(s) – Diagnostic tests, therapeutic procedures, or medical devices or technologies.

(c) Relevant – Having a significant and demonstrable bearing on the matter at hand.

(2) Pursuant to the criteria set forth in subparagraph 59G-1.010(166)(a)3., Florida Administrative Code (F.A.C.), the Agency for Health Care Administration (hereafter referred to as Agency) will determine when health services are consistent with generally accepted professional medical standards and are not experimental or investigational.

(3) Health services that are covered under the Florida Medicaid program are described in the respective coverage and limitations handbooks, policies, and fee schedules, which are incorporated by reference in the F.A.C. The public may request a health service be considered for coverage under the Florida Medicaid program by submitting a written request via e-mail to HealthServiceResearch@ahca.myflorida.com. The request must include the name, a brief description, and any additional information that supports coverage of the health service, including sources of reliable evidence as defined in paragraph 59G-1.010(84)(b), F.A.C.

(4) To determine whether the health service is consistent with generally accepted medical standards, the Agency shall consider the following factors:

(a) Evidence-based clinical practice guidelines.

(b) Published reports and articles in the authoritative medical and scientific literature related to the health service (published in peer-reviewed scientific literature generally recognized by the relevant medical community or practitioner specialty associations).

(c) Effectiveness of the health service in improving the individual's prognosis or health outcomes.

(d) Utilization trends.

(e) Coverage policies by other creditable insurance payor sources.

(f) Recommendations or assessments by clinical or technical experts on the subject or field.

(5) Based upon the information collected, a report with recommendations will be submitted to the Deputy Secretary for Medicaid (or designee) for review. The Deputy Secretary for Medicaid (or designee) will make a final determination as to whether the health service is consistent with generally accepted professional medical standards and not experimental or investigational.

(6) In order for the health service to be covered under the Florida Medicaid program, it must also meet all other medical necessity criteria as defined in subsection 59G-1.010(166), F.A.C., and funded through the General Appropriations Act or Chapter 216, F.S.

Rulemaking Authority 409.919 FS. Law Implemented 409.902, 409.906, 409.912, 409.913 FS. History–New 2-26-14, Amended 9-28-15.

Appendix Attachment

1c

ATTACHMENT C

Effects of gender affirming therapies in people with gender dysphoria: evaluation of the best available evidence

Romina Brignardello-Petersen, DDS, MSc, PhD
Wojtek Wiercioch, MSc, PhD

1.Ā Introduction

We prepared this report to fulfill a request from the Florida Agency for Health Care Administration. This report contains three documents: 1. Main document (this document) summarizing the methodology used and the findings, 2. Methods document, which provides a detailed description of the systematic methodology used to find, prioritize, appraise, and synthesize the evidence, and 3. Results document, which describes the evidence available, the estimates of the effects of gender affirming therapies, and the certainty (also known as quality) of the evidence.

This document is organized in four parts. First, we describe the credentials and expertise of the health research methodologists conducting this evidence evaluation. Second, we summarize the methodology used. Third, we summarize the main findings. Finally, we briefly discuss strengths and limitations of our process and of the evidence.

2.Ā Credentials and expertise

Two experts in health research methodology, who specialize in evidence synthesis to support decision making, prepared this report. Their relevant credentials and expertise are described below.

Dr. Romina Brignardello-Petersen: Assistant Professor at the Department of Health Research Methods, Evidence, and Impact, at McMaster University. Dr. Brignardello-Petersen obtained a DDS degree (University of Chile) in 2007, an MSc degree in Clinical Epidemiology and Health Care Research (University of Toronto) in 2012, and MSc in Biostatistics (University of Chile) in 2015, and a PhD in Clinical Epidemiology and Health Care Research (University of Toronto) in 2016. Dr. Brignardello-Petersen has worked in evidence synthesis projects since 2010, and her research has focused on the methodology for the development of Systematic Reviews and Clinical Practice Guidelines since 2012. Through January 2022, she has published 122 peer reviewed scientific articles (24 as a first author and 9 as a senior author). Dr. Brignardello-Petersen has acted as a research methodologist for several groups and organizations, including the World Health Organization, the Pan-American Health Organization, the American Society of Hematologists, the American College of Rheumatology, and the Society for Evidence Based Gender Medicine, among others. Her research program has been awarded over $2M CAD from the Canadian Institutes for Health Research. Dr. Brignardello-Petersen has no lived experience as a person or family member of a person with gender dysphoria, and her research interests are not in this area.

Dr. Wojtek Wiercioch: Postdoctoral Research Fellow at the Department of Health Research Methods, Evidence, and Impact, at McMaster University. Dr. Wiercioch obtained an MSc degree (2014, McMaster University) and a PhD degree (2020, McMaster University) in Health Research Methodology. Dr. Wiercioch has worked in evidence syntheses projects since 2011, and his research focuses on evidence synthesis, guideline development methodology, and the guideline development process. Through April

2022, he has published 86 peer-reviewed scientific articles. Dr. Wiercioch has acted as a guideline methodologist for several groups and organizations, including the World Health Organization, the American Society of Hematologists, the Endocrine Society (of America), and the American Association for Thoracic Surgeons, among others. Dr. Wiercioch has no lived experience as a person or family member of a person with gender dysphoria, and his research interests are not in this area.

3.Ā Methods

We conducted an overview of systematic reviews. We used a reproducible approach to search, select, prioritize, appraise, and synthesize the available evidence, following high methodological standards. We describe full details of the methodology in an accompanying document.

In brief, we searched for systematic reviews published in English language in Epistemonikos, OVID Medline, and grey literature sources, through April 30, 2022. We selected systematic reviews which included studies on young individuals with a diagnosis of gender dysphoria, who received puberty blockers, cross-sex hormones, or surgeries; and in which authors reported data regarding outcomes important to patients: gender dysphoria, depression, anxiety, quality of life, suicidal ideation, suicide, adverse effects, and complications. Systematic reviews could have included any type of primary study design.

The two reviewers screened all titles and abstracts, followed by full text of potentially relevant systematic reviews. We then prioritized the most useful systematic review providing evidence for each of the outcomes, using pre-established criteria that considered date of publication, applicability, availability of outcome data, methodological quality of the systematic review, and usefulness of the data synthesis conducted in the systematic review (see methods document for details).

After abstracting data from the systematic reviews, we synthesized the best available evidence for each of the outcomes, and assessed the certainty (also known as quality) of the evidence using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach. We conducted GRADE assessments using the information provided by the systematic review authors (risk of bias of primary studies, characteristics of included studies, results reported by the studies). We present the all the information about outcomes in GRADE summary of findings tables.

In addition, to evaluate the robustness of our conclusions, we systematically searched for and evaluated primary studies answering the questions of interest published after the authors of the included systematic reviews conducted their searches.

4.Ā Results

We included 61 systematic reviews, from which 3 addressed the effects of puberty blockers, 22 addressed the effects of cross-sex hormones, 30 addressed the effects of surgeries, and 6 addressed the effects of more than one of these interventions. After our prioritization exercise, we included information from 2 systematic reviews on puberty blockers, 4 on cross-sex hormones, and 8 on surgeries.

4.1 Puberty blockers

For most outcomes (except suicidality), there is no evidence about the effect of puberty blockers compared to not using puberty blockers. In other words, no studies compared the outcomes between a group of people with gender dysphoria using puberty blockers and another group of people with gender dysphoria not using them. Therefore, it is unknown whether people with gender dysphoria who use puberty blockers experience more improvement in gender dysphoria, depression, anxiety, and quality of life than those with gender dysphoria who do not use them. There is very low certainty about the effects of puberty blockers on suicidal ideation.

The studies included in the systematic review reported outcomes among a group of people with gender dysphoria after receiving puberty blockers. Low certainty evidence suggests that after treatment with puberty blockers, people with gender dysphoria experience a slight increase in gender dysphoria, and an improvement in depression, and anxiety. Low certainty evidence also suggests that a moderate percentage of patients experience adverse effects. The findings must be interpreted considering that these studies did not have a comparison group, and that it is unknown if people with gender dysphoria that do not use puberty blockers experience similar or different outcomes.

4.2 Cross sex hormones

For almost all outcomes (except breast cancer) there is no evidence about the effect of cross sex hormones compared to not using cross sex hormones. In other words, no studies compared the outcomes between a group of people with gender dysphoria using cross sex hormones and another group of people with gender dysphoria not using them. Therefore, it is unknown whether people with gender dysphoria who use cross-sex hormones experience more improvement in gender dysphoria, depression, anxiety, quality of life, and suicidality than those with gender dysphoria who do not use cross-sex hormones. There is low certainty evidence suggesting that cross-sex hormones may not increase the risk of breast cancer.

The studies included in the systematic reviews reported changes in the outcomes among a group of patients with gender dysphoria after the use of cross-sex hormones. Low certainty evidence suggests that after treatment with cross-sex hormones, people with gender dysphoria experience an improvement in gender dysphoria, depression, anxiety, and suicidality. There is very low certainty evidence about the changes in quality of life. There is moderate certainty evidence suggesting a low prevalence of venous thromboembolism after treatment with cross-sex hormones. The findings must be interpreted considering that these studies did not have a comparison group, and that it is unknown if people with gender dysphoria that do not use cross-sex hormones experience similar or different outcomes.

4.3 Surgeries

There were no systematic reviews and studies reporting on gender dysphoria, depression, anxiety, and suicidality. Therefore, the effects of surgeries on these outcomes (when compared to a group of patients with gender dysphoria who do not undergo surgery), or the changes in these outcomes (improvements or deterioration) among patients who undergo any gender-affirming surgery is unknown. Because of the lack of comparative studies, it is also unknown whether people with gender dysphoria who undergo surgeries experience more improvement in quality of life or less regret than those with gender dysphoria who do not undergo any surgeries. There is low certainty evidence suggesting that a low percentage of participants experience regret, and very low certainty evidence about changes in quality of life after surgery.

In assigned females at birth, low certainty evidence suggests that a high percentage of people are satisfied after chest surgery. There is very low certainty evidence, however, about satisfaction after bottom surgery, and about complications after both chest and bottom surgery. In assigned males at birth, low certainty evidence suggests a high percentage of people satisfied and a low percentage of people experiencing regret after vaginoplasty. There is very low certainty, however, about satisfaction with chest surgery and complications and reoperations after bottom surgery.

4.4 Evidence published after the systematic reviews selected

We found 10 relevant studies that were published after the systematic reviews were conducted. This evidence was not sufficient to importantly change the conclusions previously made.

5.Ā Discussion

5.1 Summary of the evidence

In this report, we systematically summarized the best available evidence regarding the effects of puberty blockers, cross-sex hormones, and surgeries in young people with gender dysphoria. We did not find evidence about the effect of these interventions on outcomes important to patients when compared to not receiving the intervention. We found low and very low certainty evidence suggesting improvements in gender dysphoria, depression, anxiety, and quality of life, as well as low rates of adverse events, after treatment with puberty blockers and cross-sex hormones.

5.2 Completeness and applicability

There are several gaps in the evidence regarding the effects of puberty blockers, cross-sex hormones, and surgeries in patients with gender dysphoria. Although we found some evidence for all the outcomes of interest, the evidence is suboptimal: several limitations included the lack of studies with a comparison group, and the risk of bias and imprecision, resulting in low or very low certainty evidence for all outcomes.

The applicability of the evidence may also be limited. Although we only rated down for indirectness when it was considered a serious problem (i.e., in evidence about the effects of surgeries, which was collected from people who were importantly older than the target population in this report), there are also potential applicability issues to consider in the evidence regarding the effects of puberty blockers and cross-sex hormones. It is not clear to what extent the people included in the studies were similar enough to the people seeking these treatment options today. For example, some of the included studies were conducted in people who had a diagnosis of gender dysphoria confirmed with strict criteria, as well as a supportive environment. It is important to take into account to what extent this may compromise the applicability of the results to people who are not in the same situation.

5.3 Strengths and limitations of the process for developing this report

We followed a reproducible process for developing this report. We used the highest methodological standards and the approach to evidence synthesis we generally use when supporting organizations in the development of their guidelines. This approach is based on prioritizing the sources of evidence most likely to be informative (i.e., to identify and use the evidence with the highest certainty level).

To follow the principles for evidence-based decision making, which require using the best available evidence to inform decisions, we summarized the best available evidence. Because knowing the best

available evidence necessitates being aware of all the available evidence, we based this report on systematic reviews of the literature. We chose the most trustworthy and relevant systematic reviews among many published reviews.

One potential limitation of the process is that, due to feasibility concerns, we relied on the information reported by the systematic reviewers. Most of the systematic reviews we used, unfortunately, were judged at moderate or low methodological quality, which may raise concerns about the trustworthiness of the evidence presented in this report. We believe, however, that the results and conclusions of this report would not be importantly different had the systematic reviews been conducted following higher methodological standards. Because there are no randomized controlled trials, well-conducted comparative observational studies, or very large case series (which include a large sample of consecutive patients who are representative of the whole population) addressing the effects of puberty blockers, cross-sex hormones, and surgeries; the certainty of the evidence about the effects of these interventions is likely to continue being low or very low, even if a few more studies are included (as observed after searching for primary studies published after the reviews were conducted) or some data points were reported inaccurately in the systematic reviews.

Also due to feasibility concerns, the scope of this report was limited to outcomes that are important to patients. Although some may question the decision of not including surrogate outcomes for which there is evidence available (e.g. bone density, blood pressure), decision makers should rarely consider these outcomes and should instead focus on outcomes that do matter to people and stakeholders (e.g., fractures, cardiovascular events).

5.4 Implications

The evidence evaluating the effects of puberty blockers, cross-sex hormones, and surgeries in people with gender dysphoria has important limitations. Therefore, decisions regarding their use should carefully consider other relevant factors. At a patient level, these factors include patients' values and preferences (how patients trade off the potential benefit and harms - what outcomes are more important to them), and resources needed to provide the interventions (and the availability of such resources). At a population level, in addition to these factors, it would be important to consider resources needed to implement the interventions, feasibility, acceptability by relevant stakeholders, and equity.

It is important to note that when there is low or very low certainty evidence, it is rarely appropriate to make decisions that will be applied to the majority of the patients (equivalent to strong recommendations). This implies, at the patient level, that shared decision making is a key part of the decision-making process. At a policy level, extensive debate may be needed.

6.Ā Conclusions

Due to the important limitations in the body of evidence, there is great uncertainty about the effects of puberty blockers, cross-sex hormones, and surgeries in young people with gender dysphoria. This evidence alone is not sufficient to support whether using or not using these treatments. We encourage decision makers to be explicit and transparent about which factors play an important role in their decision, and how they are weighed and traded off.

Methods

To ensure completeness and feasibility of the evidence review, we used an approach in which we prioritized the types of studies according to the design that was more likely to provide the best available evidence. First, we searched for systematic reviews of the literature. Second, we appraised all existing systematic reviews to select the most trustworthy (highest methodological quality, most up-to-date, most applicable) from which to draw conclusions. Third, we used the information presented in the systematic reviews to abstract information regarding the effects of the interventions of interest. Fourth, we assessed the certainty of the evidence (also known as quality of the evidence) abstracted from the selected systematic reviews. We planned to search for primary studies if systematic reviews were not found.

Information sources: We searched for existing systematic reviews in:

1. Epistemonikos (https://www.epistemonikos.org), an electronic database that focuses on systematic reviews. We used a comprehensive search strategy based on the population, using the terms "gender dysphoria", "gender identity disorder" and "transgender". We conducted this search on April 23, 2022.
2. OVID Medline. We used a search strategy based on the population and the interventions of interest, as well as an adaptation of a filter for systematic reviews from the Health Information Research Unit at McMaster University. We conducted this search on April 23, 2022.
3. Grey literature: we conducted a manual search in the websites of specific health agencies: National Institutes for Health and Care Excellence (NICE), Agency for Healthcare Research and Quality (AHRQ), Canada's Drug and Health Technology Agency (CADTH), and the website from the Society for Evidence-Based Gender Medicine (SEGM). We conducted these searches between April 27-30, 2022.

We used no date limits for the searches, but we did limit to systematic reviews published in English. Search strategies are available in Appendix 1.

Eligibility criteria: We included systematic reviews, which we defined as:

1. Reviews in which the authors searched for studies to include in at least one electronic database, and in which there were eligibility criteria for including studies and a methodology for assessing and synthesizing the evidence, or
2. Reviews in which the authors searched for studies to include in at least one electronic database, and although there was no description of eligibility criteria or methodology, the presentation of the results strongly suggested that the authors used systematic methods (e.g. flow chart depicting study selection, tables with the same information from all included studies, synthesis of data at the outcome level).

We screened systematic reviews using the following criteria for inclusion:

- ***Type of participants:*** Young individuals (< 25 years old) with a diagnosis of gender dysphoria/gender identity disorder. We included reviews in which authors used any label and diagnostic criteria for this condition. We included reviews in which the participants in the reported studies were older if it was the only evidence available for a specific question. We excluded reviews with mixed populations (i.e. with and without gender dysphoria) in which people without gender dysphoria constituted more than 20% of the total sample.

- ***Type of Interventions*:** Puberty blockers, cross-sex hormones, gender affirming surgeries. We included any type of puberty blockers and cross-sex hormones, provided with any regimen. We included the following surgeries: phalloplasty, vaginoplasty, and chest surgery (mastectomy or breast implants/augmentation). We only included these when they were performed for the first time (i.e., not revision surgeries).

- ***Type of comparison:*** When the systematic reviews included comparative studies, the comparator of interest was no intervention. Participants could have received psychotherapy or counselling as a cointervention (in both groups).

- ***Type of outcomes:*** Gender dysphoria, mental health outcomes (depression and anxiety), quality of life, suicidal ideation, suicide, adverse effects (for puberty blockers and cross-sex hormones only), and satisfaction, complications, reoperation, and regret (for surgeries only). We included any length of follow-up. We excluded surrogate outcomes such as blood pressure, bone mineral density, kidney or liver function test values, etc.

- ***Type of studies included in the systematic reviews:*** Any clinical study (studies in which the researchers recruited and measured outcomes in humans) regardless of study design. This included randomized clinical trials, comparative observational studies, and case series. Because we could not quantify effect measures, incidence, or prevalence, we excluded case reports.

We excluded systematic reviews published only in abstract format, and those that we could not retrieve in full text (no access through the McMaster University library, or open access online).

Selection process: The two reviewers screened all titles and abstracts independently and in duplicate, followed by screening of full texts of potentially eligible systematic reviews independently and in duplicate, using the systematic review online application Covidence (https://www.covidence.org). We solved disagreements by consensus.

To select the most useful systematic reviews among all of those that met the eligibility criteria, we used the following prioritization criteria:

1. Date of publication: we prioritized systematic reviews published within the last 3 years (2020-2022)

2. Match between eligibility criteria of the review and the question of interest: we prioritized reviews in which the authors specifically included the population, intervention, comparison, and outcomes of interest for this evidence review
3. Outcome data available: we prioritized systematic reviews in which the authors report outcome data
4. Methodological quality: we used a modified version of the items in AMSTAR 2.[1] We modified the items to ensure assessment of methodological rather than reporting quality (Table 1). We rated each systematic review as having high, moderate, low, or critically low methodological quality, according to the guidance from the developers of the tool.[1] We reached consensus on critical items that determined this rating (Table 1). We prioritized selection of systematic reviews with highest methodological quality.

For surgical interventions, in addition, we prioritized systematic reviews that covered all gender affirming surgeries (instead of focusing on a specific type of surgery).

We selected a systematic review specifically for each of the outcomes of interest. In other words, we chose the best systematic review to inform each outcome. Each systematic review, however, could inform more than one outcome.

Table 1: Items used to rate the methodological quality of the eligible systematic reviews

AMSTAR Item	Modification to measure methodological quality
1. Did the research questions and inclusion criteria for the review include the components of PICO?	Does the review have a clear question and are the eligibility criteria for studies consistent with the question?
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	No modification needed
3. Did the review authors explain their selection of the study designs for inclusion in the review?	No modification needed
4. Did the review authors use a comprehensive literature search strategy?	Did the authors search in at least 2 electronic databases, using a reproducible search strategy?
5. Did the review authors perform study selection in duplicate?	No modification needed
6. Did the review authors perform data extraction in duplicate?	No modification needed
7. Did the review authors provide a list of excluded studies and justify the exclusions?	No modification needed
8. Did the review authors describe the included studies in adequate detail?	No modification needed
9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	No modification needed

10. Did the review authors report on the sources of funding for the studies included in the review?	Did the review authors consider conflicts of interest and how they may have affected the results of the primary studies?
11. If meta-analysis was performed, did the review authors use appropriate methods for statistical combination of results?	Was the synthesis of evidence done appropriately? (outcome level, appropriate meta analysis or narrative synthesis)
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	Did authors use subgroup or sensitivity analysis to assess the effect of risk of bias in meta-analytic results? Likely not applicable to most cases
13. Did the review authors account for RoB in primary studies when interpreting/discussing the results of the review?	Did the review authors incorporate an assessment of risk of bias at the outcome level when drawing conclusions?
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	Did the review authors incorporate an assessment of heterogeneity at the outcome level when drawing conclusions?
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	Did the authors address publication bias? (regardless of whether synthesis was using a meta-analysis or narrative)
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	Did the authors report conflicts of interest and did they manage any existing conflict of interest appropriately?

Shaded items were items considered critical.

Data abstraction: We abstracted outcome data from each of the systematic reviews. To ensure feasibility, we used the data as reported by the authors of the review and did not re-abstract data from the primary studies. One reviewer abstracted the data and a second reviewer checked the data for accuracy.

Data synthesis: Using the systematic reviews prioritized, we synthesized the evidence at the outcome level. Because of the higher likelihood of it resulting in higher certainty of evidence (details below) for each outcome, when there was comparative data (i.e. comparison of outcomes between an untreated and a treated group) and non-comparative data (i.e. changes from before to after treatment in one group, or only outcomes after treatment), we prioritized comparative data.

We prioritized numerical results (i.e. magnitudes of effect) and reported estimates and their 95% confidence intervals (CIs). When results were not reported in that way, we calculated the estimates and CIs when systematic review authors provided sufficient information. When necessary, we assumed moderate correlation coefficients for the changes between baseline and follow up (coefficient= 0.4). When this information was not available we reported narratively the effect estimates and ranges.

When a specific study reported the same outcome measured by more than one scale, we chose the scale presented first. We highlighted situations when the results obtained with other scales were importantly different.

When the same outcome was reported by more than one study but we could not pool the results, we created narrative syntheses.

Certainty of evidence: For each outcome, we assessed the certainty of the evidence (also known as quality of the evidence) using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach.[2] The certainty of evidence can be rated as high, moderate, low, or very low (Table 2). For effects of interventions, the certainty of the evidence started as high and could be rated down due to serious concerns about risk of bias, inconsistency, indirectness, imprecision, and publication bias. For inferences about the effect of using a treatment versus no treatment, when there was no comparison group, we assessed risk of bias as very serious and rated down the certainty of the evidence 2 levels by default. We used the same principles when assessing the certainty of the evidence in estimates of prevalence or rates, but did not judge risk of bias as resulting in very serious concerns due to lack of a comparison group. For all assessments, we used the information presented by the authors of the systematic review (e.g. assessments of risk of bias of the included studies, effect estimates from studies).

Table 2: GRADE levels of certainty of the evidence

Certainty level	Definition
High ⊕⊕⊕⊕	We are very confident that the true result (effect estimate/ prevalence/ mean, etc.) lies close to that of the estimate of the result
Moderate ⊕⊕⊕◯	We are moderately confident in the result: the true result is likely to be close to the estimate of the result, bur there is a possibility that it is substantially different
Low ⊕⊕◯◯	Our confidence in the result is limited: the true result may be substantially different from the estimate of the result
Very low ⊕◯◯◯	We have very little confidence in the result: the true result is likely to be substantially different from the estimate of the result

Presentation of results: We created GRADE Summary of Findings tables in which we describe the evidence available for each of the outcomes, and the certainty of the evidence. These tables contain the following information:

- Outcomes: measurement method (including scales, if applicable) and follow-up
- Estimates of effect: absolute and relative estimates of effect, and their corresponding 95% CIs.
- Number of studies and participants providing evidence for the outcome
- GRADE certainty of the evidence, with a link to detailed explanations (provided at the bottom of the table) of why the certainty of the evidence was rated at a specific level
- A narrative statement about what happens with the outcome, based on the estimate of effect and certainty of evidence.

Searching for new evidence not included in the systematic reviews: To assess if newer evidence not included in the included systematic reviews would change the conclusions importantly, we searched for and assessed primary studies answering the questions of interest that were published after the authors of such systematic reviews conducted their searches. We defined an important change in conclusions as a change in the certainty of the evidence (from low/ very low/ not available to high/ moderate).

We searched OVID Medline from January 1, 2019 through May 12, 2022, for studies published in English. We included studies if they enrolled young individuals (< 25 years old, with at least 20% of the people being this age) with a diagnosis of gender dysphoria/gender identity disorder, who received puberty blockers, cross-sex hormones, or surgeries; and measured any of the outcomes of interest.

For outcomes that should be evaluated in a comparative manner (e.g., depression, anxiety, etc.), because they are the only type of study design that would change the conclusions importantly, we selected comparative clinical studies (studies in which the researchers recruited and measured outcomes in humans, and compared a group of people who received the intervention with another one who did not receive the intervention). This included randomized clinical trials, and comparative observational studies. For outcomes that can only occur when the treatment is administered, we included non-comparative observational studies (case series). For these to change conclusions, they should have a sufficiently large sample size, and therefore we excluded case series in which the researchers reported information from <100 people.

Two reviewers screened the potentially relevant articles at title and abstract and full text screening stage. We abstracted relevant study characteristics and outcome data, and assessed risk of bias of comparative studies using the most relevant domains of the Risk of Bias for non-Randomized studies of Interventions (ROBINS-I) tool[3] (table 3). For non-comparative studies, we used a list of custom items that captured the most important potential risk of bias concerns of case series (table 4). We judged the risk of bias of each study as the highest risk of bias of any of the domains assessed (e.g., one domain judged at critical risk of bias resulted in the study judged at critical risk of bias). We summarized this information at the study and judged whether it would have changed the conclusions importantly if added to the body of evidence from the systematic reviews.

Table 3: Domains used to assess risk of bias of comparative studies

Domain	Low	Critical
Confounding	Adjusted for all relevant confounding factors	No adjustment
Classification of intervention	Intervention recorded prospectively or from medical records	Asked patients to recall whether they received the intervention
Deviation from intended interventions	No cointerventions or cointerventions balanced between the groups	Cointerventions unbalanced between the groups

Missing data	More than 90% of patients who started the study provided outcome data	Less thank 50% of patients who started the study provided outcome data
Measurement of outcome	All outcomes measured in the same way in both groups	Outcomes measured differently in both groups

Each domain could be judged at low, moderate, serious, or critical risk of bias. In addition, information could be insufficient to make a judgment. The table describes the criteria used to judge a domain in the extreme categories.

Table 4: Domains used to assess risk of bias of non- comparative studies

Domain	Low	High
Representativeness of the sample	Included all consecutive patients	Highly selected sample based on specific characteristics related with the prognosis after treatment
Classification of the intervention	Intervention recorded prospectively or from medical records	Asked patients to recall whether they received the intervention
Deviation from intended interventions	No cointerventions outside what would be observed in practice (or in a small proportion of patients)	Most patients received co interventions that could influence the outcomes
Missing data	More than 90% of patients who started the study provided outcome data	Less thank 50% of patients who started the study provided outcome data
Measurement of outcome	Outcomes measured prospectively or from medical records	Outcomes reported by the patients and/ or needed to recall what happened a long time ago

Each domain could be judged at low, moderate, or high risk of bias. In addition, information could be insufficient to make a judgment. The table describes the criteria used to judge a domain in the extreme categories.

References

1. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *Bmj* 2017;358:j4008. doi: 10.1136/bmj.j4008 [published Online First: 2017/09/25]
2. Blashem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of the evidence. *Journal of clinical epidemiology* 2011;64:401-06.
3. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ (Clinical research ed)* 2016;355:i4919. doi: 10.1136/bmj.i4919 [published Online First: 2016/10/14]

Search Strategies

Questions Covered:

PICO questions:

1. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of treatment with **puberty blockers (gonadotrophin releasing hormone (GnRH) analogues)** compared to no puberty blockers?

2. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of treatment with **cross-sex hormones** compared to no cross-sex hormones?

3. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of **gender-affirming surgeries** compared to no surgery?

Search Strategies:

Note: Population, puberty blocker, cross-sex hormones search blocks adapted from NICE (2020) evidence reviews. Gender-affirming search block adapted from Wernick *et al.* 2019. Systematic reviews filter adapted from McMaster University Health Information Research Unit (HIRU).

Databases: Medline, Epistemonikos
Grey Literature: CADTH, AHRQ, SEGM, NICE

Medline

OVERVIEW	
Interface:	Ovid
Databases:	OVID Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
Study Types:	Systematic Reviews
Search Run:	April 23, 2022

Search Strategy: search terms [number of results]

Population

1	exp "Sexual and Gender Minorities"/	12385
2	Gender Dysphoria/	774
3	Gender Identity/	20481
4	Gender Role/	197
5	"Sexual and Gender Disorders"/	81
6	Transsexualism/	4236
7	Transgender Persons/	5303
8	Health Services for Transgender Persons/	186

9 exp Sex Reassignment Procedures/ 1208
10 gender identity disorder.mp. 492
11 non-binary.mp. 566
12 transgender.mp. 9989
13 (gender* adj3 (dysphori* or disorder* or distress or nonconform* or non-conform* or atypical or incongru* or identi* or disorder* or confus* or minorit* or queer* or variant or diverse or creative or explor* or question* or expan* or fluid)).tw. 16428
14 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition* or expression*)).tw. 13749
15 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. 19665
16 (genderfluid or genderqueer or agender).mp. 130
17 ((correct or chosen) adj3 name).mp. 591
18 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. 135313
19 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition* or expression*)).tw. 13749
20 (male-to-female or m2f or female-to-male or f2m).tw. 148579
21 or/1-20 342948

Cross-Sex Hormones

22 Hormones/ad, tu, th 4676
23 exp Progesterone/ad, tu, th 11265
24 exp Estrogens/ad, tu, th 29635
25 exp Gonadal Steroid Hormones/ad, tu, th 35375
26 (progesteron* or oestrogen* or estrogen*).tw. 223307
27 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. 1488
28 exp Estradiol/ad, tu, th 11197
29 exp Testosterone/ad, tu, th 8710
30 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. 86509
31 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. 100252
32 or/22-31 345895

Puberty Blockers

33 Gonadotropin-Releasing Hormone/ 28809
34 (pubert* adj3 block*).ti,ab. 141
35 ((gonadotrophin or gonadotropin) and releasing).ti,ab. 20121
36 (GnRH adj2 analog*).ti,ab. 2878
37 GnRH*.ti,ab. 24390
38 "GnRH agonist*".ti,ab. 4749
39 Triptorelin Pamoate/ 1981
40 triptorelin.ti,ab. 821
41 arvekap.ti,ab. 1

42 ("AY 25650" or AY25650).ti,ab. 1
43 ("BIM 21003" or BIM21003).ti,ab. 0
44 ("BN 52014" or BN52014).ti,ab. 0
45 ("CL 118532" or CL118532).ti,ab. 0
46 Debio.ti,ab. 119
47 diphereline.ti,ab. 28
48 moapar.ti,ab. 0
49 pamorelin.ti,ab. 1
50 trelstar.ti,ab. 3
51 triptodur.ti,ab. 1
52 ("WY 42422" or WY42422).ti,ab. 0
53 ("WY 42462" or WY42462).ti,ab. 0
54 gonapeptyl.ti,ab. 0
55 decapeptyl.ti,ab. 225
56 salvacyl.ti,ab. 0
57 Buserelin/ 2137
58 buserelin.ti,ab. 1395
59 onist.ti,ab. 0
60 ("hoe 766" or hoe-766 or hoe766).ti,ab. 72
61 profact.ti,ab. 2
62 receptal.ti,ab. 31
63 suprecur.ti,ab. 5
64 suprefact.ti,ab. 25
65 tiloryth.ti,ab. 0
66 histrelin.ti,ab. 78
67 "LHRH-hydrogel implant".ti,ab. 1
68 ("RL 0903" or RL0903).ti,ab. 1
69 ("SPD 424" or SPD424).ti,ab. 1
70 goserelin.ti,ab. 1016
71 Goserelin/ 1643
72 ("ici 118630" or ici118630).ti,ab. 51
73 ("ZD-9393" or ZD9393).ti,ab. 0
74 zoladex.ti,ab. 388
75 leuprorelin.ti,ab. 525
76 carcinil.ti,ab. 0
77 enanton*.ti,ab. 26
78 ginecrin.ti,ab. 0
79 leuplin.ti,ab. 15
80 Leuprolide/ 3018
81 leuprolide.ti,ab. 2004
82 lucrin.ti,ab. 16
83 lupron.ti,ab. 183
84 provren.ti,ab. 0
85 procrin.ti,ab. 3
86 ("tap 144" or tap144).ti,ab. 41
87 (a-43818 or a43818).ti,ab. 3
88 Trenantone.ti,ab. 2
89 staladex.ti,ab. 0

90 prostap.ti,ab. 6
91 Nafarelin/ 327
92 nafarelin.ti,ab. 263
93 ("76932-56-4" or "76932564").ti,ab. 0
94 ("76932-60-0" or "76932600").ti,ab. 0
95 ("86220-42-0" or "86220420").ti,ab. 0
96 ("rs 94991 298" or rs94991298).ti,ab. 0
97 synarel.ti,ab. 13
98 deslorelin.ti,ab. 306
99 gonadorelin.ti,ab. 237
100 ("33515-09-2" or "33515092").ti,ab. 0
101 ("51952-41-1" or "51952411").ti,ab. 0
102 ("52699-48-6" or "52699486").ti,ab. 0
103 cetrorelix.ti,ab. 520
104 cetrotide.ti,ab. 52
105 ("NS 75A" or NS75A).ti,ab. 0
106 ("NS 75B" or NS75B).ti,ab. 0
107 ("SB 075" or SB075).ti,ab. 1
108 ("SB 75" or SB75).ti,ab. 67
109 gonadoliberin.ti,ab. 151
110 kryptocur.ti,ab. 7
111 cetrorelix.ti,ab. 520
112 cetrotide.ti,ab. 52
113 antagon.ti,ab. 18
114 ganirelix.ti,ab. 160
115 ("ORG 37462" or ORG37462).ti,ab. 3
116 orgalutran.ti,ab. 26
117 ("RS 26306" or RS26306).ti,ab. 5
118 ("AY 24031" or AY24031).ti,ab. 0
119 factrel.ti,ab. 13
120 fertagyl.ti,ab. 12
121 lutrelef.ti,ab. 5
122 lutrepulse.ti,ab. 3
123 relefact.ti,ab. 10
124 fertiral.ti,ab. 0
125 (hoe471 or "hoe 471").ti,ab. 6
126 relisorm.ti,ab. 4
127 cystorelin.ti,ab. 19
128 dirigestran.ti,ab. 5
129 or/33-128 47108

Gender-affirming Surgeries

130 virilization/ 2309
131 (virilism or virili?ation or masculini?ation).mp. 5657
132 feminization/ 797
133 femini?ation.mp. 3420
134 (vaginoplasty or vaginoplasties).mp. 1022

135 exp Vagina/ or *Reconstructive Surgical Procedures/ 78841
136 (vaginoplasty or vaginoplasties).mp. 1022
137 (phalloplasty or phalloplasties).mp. 561
138 exp Penile Prosthesis/ 1636
139 "penile reconstruction".mp. 292
140 (vagina reconstruction or vaginal reconstruction).mp. 549
141 (genitoplasty or genitoplasties).mp. 263
142 transsexualism/su [Surgery] 1007
143 sex reassignment.mp. 1668
144 sex transformation.mp. 42
145 or/130-144 91560

Systematic Review Filter

147 meta-analysis/ 158633
148 (meta anal* or meta-anal* or metaanal*).ti,ab. 231876
149 ((systematic or evidence) adj2 (review* or overview*)).ti,ab. 279806
150 ((pool* or combined) adj2 (data or trials or studies or results)).ab. 65411
151 (search strategy or search criteria or systematic search or study selection or data extraction).ab. 70886
152 (search* adj4 literature).ab. 84593
153 or/146-152 521554

Combine Interventions and Population

154 32 or 129 or 145 459771
155 21 and 154 17838

Limit to Systematic Reviews in English Language

156 153 and 155 295
157 limit 156 to english language 288

Epistemonikos

OVERVIEW	
Interface:	https://www.epistemonikos.org/
Database:	Epistemonikos
Study Types:	Systematic Reviews
Search Run:	April 23, 2022

Search Strategy: search terms [number of results]

Population

(title:((title:(gender dysphoria) OR abstract:(gender dysphoria)) OR (title:(gender identity disorder) OR abstract:(gender identity disorder)) OR (title:(transgender) OR abstract:(transgender))) OR abstract:((title:(gender dysphoria) OR abstract:(gender dysphoria)) OR (title:(gender identity disorder) OR abstract:(gender identity disorder)) OR (title:(transgender) OR abstract:(transgender))))

Limit to Systematic Reviews

*Limited by publication type "systematic review" [425]

Canadian Agency for Drugs and Technologies in Health (CADTH)

OVERVIEW	
Interface:	https://www.cadth.ca/
Database:	CADTH
Study Types:	Systematic Reviews, Health Technology Reviews
Search Run:	April 27, 2022

Search Strategy: search terms [number of results]

"gender dysphoria" [10]
Limit to Health Technology Review [2]

"transgender" [9]
Limit to Health Technology Review [5]

"gender identity disorder" [1]

Agency for Healthcare Research and Quality (AHRQ)

OVERVIEW	
Interface:	https://search.ahrq.gov/
Database:	AHRQ
Study Types:	Evidence Based Practice (EPC) Centre Reports, Full Research Reports, Health Technology Assessments
Search Run:	April 29, 2022

Search Strategy: search terms [number of results]

Search titles only: "gender identity disorder" "gender dysphoria" "transgender" [7]

Society for Evidence-based Gender Medicine (SEGM)

OVERVIEW	
Interface:	https://segm.org/news
Database:	SEGM News
Study Types:	Systematic Reviews
Search Run:	April 30, 2022

Search Strategy: search terms [number of results]

Find in page: "systematic" [5]

National Institute for Health and Care Excellence (NICE)

OVERVIEW	
Interface:	https://www.nice.org.uk/
Database:	NICE
Study Types:	Systematic Reviews, Guidelines with Systematic Reviews
Search Run:	April 30, 2022

Search Strategy: search terms [number of results]

gender dysphoria [1]
Limit to Guidance [1]

transgender [10]
Limit to Guidance [7]

gender identity disorder [9]
Limit to Guidance [8]

Search Strategies – Individual Studies

Questions Covered:

PICO questions:

1. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of treatment with **puberty blockers (gonadotrophin releasing hormone (GnRH) analogues)** compared to no puberty blockers?

2. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of treatment with **cross-sex hormones** compared to no cross-sex hormones?

3. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of **gender-affirming surgeries** compared to no surgery?

Search Strategies:

Note: Population, puberty blocker, cross-sex hormones search blocks adapted from NICE (2020) evidence reviews. Gender-affirming search block adapted from Wernick *et al.* 2019.

Databases: Medline

Medline

OVERVIEW	
Interface:	Ovid
Databases:	OVID Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
Study Types:	Any
Search Run:	May 12, 2022

Search Strategy: search terms [number of results]

Population

1	exp "Sexual and Gender Minorities"/	12631
2	Gender Dysphoria/	781
3	Gender Identity/	20586
4	Gender Role/	204
5	"Sexual and Gender Disorders"/	81
6	Transsexualism/	4259
7	Transgender Persons/	5371
8	Health Services for Transgender Persons/	187
9	exp Sex Reassignment Procedures/	1211
10	gender identity disorder.mp.	492

11 non-binary.mp. 574
12 transgender.mp. 10079
13 (gender* adj3 (dysphori* or disorder* or distress or nonconform* or non-conform* or atypical or incongru* or identi* or disorder* or confus* or minorit* or queer* or variant or diverse or creative or explor* or question* or expan* or fluid)).ti,ab. 16546
14 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).ti,ab. 9375
15 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).ti,ab. 19788
16 (genderfluid or genderqueer or agender).mp. 132
17 ((correct or chosen) adj3 name).mp. 591
18 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).ti,ab. 135744
19 (male-to-female or m2f or female-to-male or f2m).ti,ab. 149067
20 or/1-19 341083

Cross-sex Hormones

21 Hormones/ad, tu, th 4690
22 exp Progesterone/ad, tu, th 11270
23 exp Estrogens/ad, tu, th 29646
24 exp Gonadal Steroid Hormones/ad, tu, th 35401
25 (progesteron* or oestrogen* or estrogen*).ti,ab. 223689
26 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).ti,ab. 1507
27 exp Estradiol/ad, tu, th 11200
28 exp Testosterone/ad, tu, th 8722
29 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).ti,ab. 86670
30 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).ti,ab. 100411
31 or/21-30 346508

Puberty Blockers

32 Gonadotropin-Releasing Hormone/ 28845
33 (pubert* adj3 block*).ti,ab. 142
34 ((gonadotrophin or gonadotropin) and releasing).ti,ab. 20158
35 (GnRH adj2 analog*).ti,ab. 2879
36 GnRH*.ti,ab. 24437
37 "GnRH agonist*".ti,ab. 4763
38 Triptorelin Pamoate/ 1983
39 triptorelin.ti,ab. 822
40 arvekap.ti,ab. 1
41 ("AY 25650" or AY25650).ti,ab. 1
42 ("BIM 21003" or BIM21003).ti,ab. 0
43 ("BN 52014" or BN52014).ti,ab. 0
44 ("CL 118532" or CL118532).ti,ab. 0

45 Debio.ti,ab. 119
46 diphereline.ti,ab. 28
47 moapar.ti,ab. 0
48 pamorelin.ti,ab. 1
49 trelstar.ti,ab. 3
50 triptodur.ti,ab. 1
51 ("WY 42422" or WY42422).ti,ab. 0
52 ("WY 42462" or WY42462).ti,ab. 0
53 gonapeptyl.ti,ab. 0
54 decapeptyl.ti,ab. 225
55 salvacyl.ti,ab. 0
56 Buserelin/ 2137
57 buserelin.ti,ab. 1396
58 onist.ti,ab. 0
59 ("hoe 766" or hoe-766 or hoe766).ti,ab. 72
60 profact.ti,ab. 2
61 receptal.ti,ab. 31
62 suprecur.ti,ab. 5
63 suprefact.ti,ab. 25
64 tiloryth.ti,ab. 0
65 histrelin.ti,ab. 78
66 "LHRH-hydrogel implant".ti,ab. 1
67 ("RL 0903" or RL0903).ti,ab. 1
68 ("SPD 424" or SPD424).ti,ab. 1
69 goserelin.ti,ab. 1017
70 Goserelin/ 1644
71 ("ici 118630" or ici118630).ti,ab. 51
72 ("ZD-9393" or ZD9393).ti,ab. 0
73 zoladex.ti,ab. 388
74 leuprorelin.ti,ab. 529
75 carcinil.ti,ab. 0
76 enanton*.ti,ab. 26
77 ginecrin.ti,ab. 0
78 leuplin.ti,ab. 15
79 Leuprolide/ 3018
80 leuprolide.ti,ab. 2003
81 lucrin.ti,ab. 16
82 lupron.ti,ab. 183
83 provren.ti,ab. 0
84 procrin.ti,ab. 3
85 ("tap 144" or tap144).ti,ab. 41
86 (a-43818 or a43818).ti,ab. 3
87 Trenantone.ti,ab. 2
88 staladex.ti,ab. 0
89 prostap.ti,ab. 6
90 Nafarelin/ 327
91 nafarelin.ti,ab. 263
92 ("76932-56-4" or "76932564").ti,ab. 0

93 ("76932-60-0" or "76932600").ti,ab. 0
94 ("86220-42-0" or "86220420").ti,ab. 0
95 ("rs 94991 298" or rs94991298).ti,ab. 0
96 synarel.ti,ab. 13
97 deslorelin.ti,ab. 310
98 gonadorelin.ti,ab. 238
99 ("33515-09-2" or "33515092").ti,ab. 0
100 ("51952-41-1" or "51952411").ti,ab. 0
101 ("52699-48-6" or "52699486").ti,ab. 0
102 cetrorelix.ti,ab. 520
103 cetrotide.ti,ab. 52
104 ("NS 75A" or NS75A).ti,ab. 0
105 ("NS 75B" or NS75B).ti,ab. 0
106 ("SB 075" or SB075).ti,ab. 1
107 ("SB 75" or SB75).ti,ab. 67
108 gonadoliberin.ti,ab. 152
109 kryptocur.ti,ab. 7
110 cetrorelix.ti,ab. 520
111 cetrotide.ti,ab. 52
112 antagon.ti,ab. 18
113 ganirelix.ti,ab. 161
114 ("ORG 37462" or ORG37462).ti,ab. 3
115 orgalutran.ti,ab. 26
116 ("RS 26306" or RS26306).ti,ab. 5
117 ("AY 24031" or AY24031).ti,ab. 0
118 factrel.ti,ab. 13
119 fertagyl.ti,ab. 12
120 lutrelef.ti,ab. 5
121 lutrepulse.ti,ab. 3
122 relefact.ti,ab. 10
123 fertiral.ti,ab. 0
124 (hoe471 or "hoe 471").ti,ab. 6
125 relisorm.ti,ab. 4
126 cystorelin.ti,ab. 19
127 dirigestran.ti,ab. 5
128 or/32-127 47179

Surgery

129 virilization/ 2309
130 (virilism or virili?ation or masculini?ation).mp. 5664
131 feminization/ 798
132 femini?ation.mp. 3425
133 (vaginoplasty or vaginoplasties).mp. 1032
134 (vaginoplasty or vaginoplasties).mp. 1032
135 (phalloplasty or phalloplasties).mp. 561
136 exp Penile Prosthesis/ 1642
137 "penile reconstruction".mp. 292

138 (vagina reconstruction or vaginal reconstruction).mp. 550
139 (genitoplasty or genitoplasties).mp. 263
140 transsexualism/su [Surgery] 1007
141 sex reassignment.mp. 1674
142 sex transformation.mp. 42
143 or/129-142 14290

Any intervention AND population

144 31 or 128 or 143 386835
145 20 and 144 16516

Limit to Humans

146 animals/ not humans/ 4972586
147 145 not 146 9281
148 limit 147 to humans 7901

Limit to Publication Year 2019 to Current

149 limit 148 to yr="2019 -Current" 1859

Results

Search results and eligible reviews: After screening 647 records found through our searches, we found 61 eligible systematic reviews. From these, 27 were published between 2020 and 2022 (Figure 1). Overall, 4% (1/27) of the reviews were judged to be of high methodological quality, 15% (4/27) were moderate methodological quality, 37% (10/27) were low methodological quality, and 44% (12/27) were critically low methodological quality.

We provide reasons for excluding systematic reviews in appendix 1.

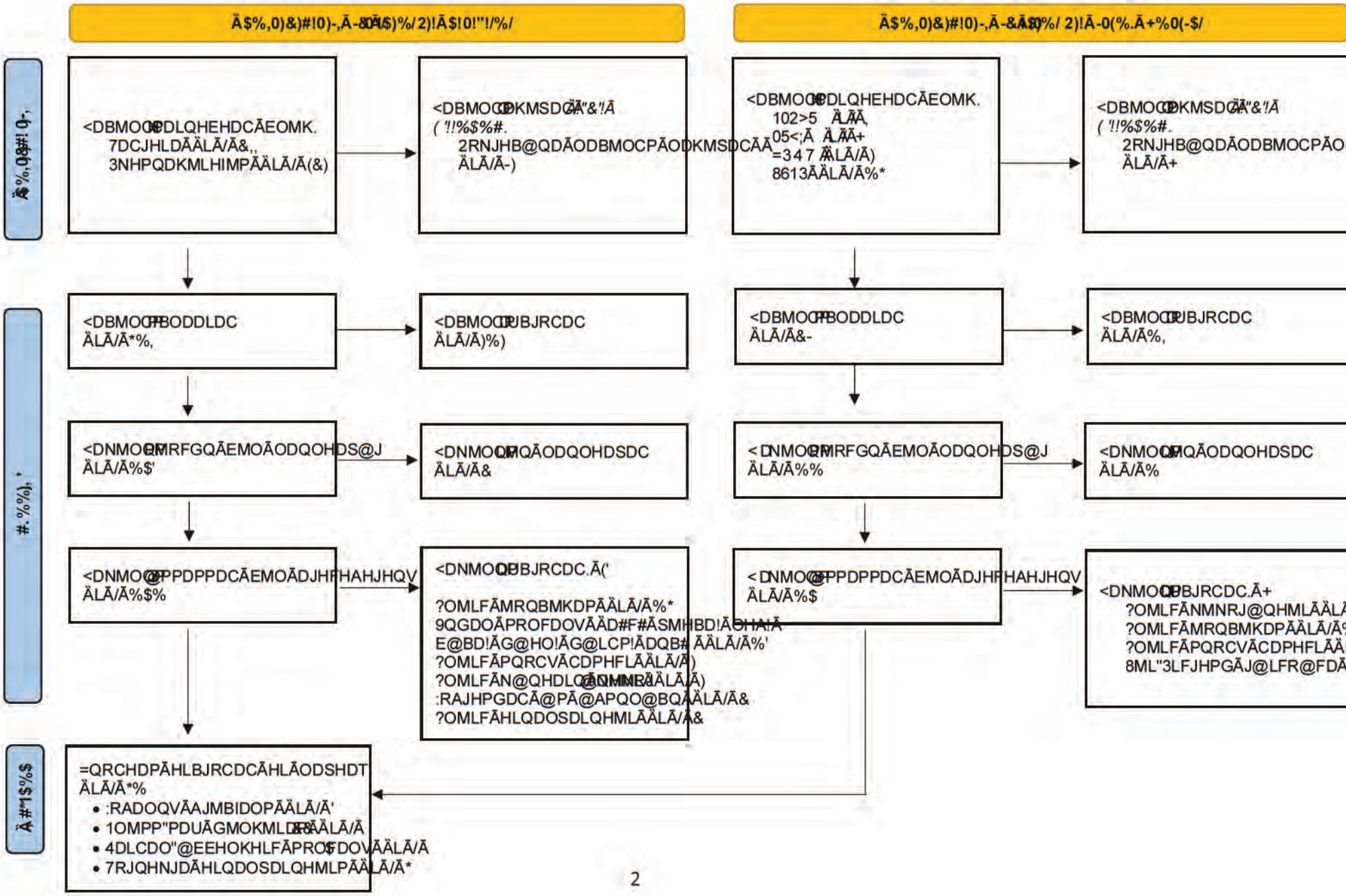
Ã$%,0)&)#!0)-,Ã-8Ã$)%/2)!Ã$!0!"!/%/
Ã$%,0)&)#!0)-,Ã-&Ã$)%/2)!Ã-0(%.Ã+%0(-$/
#.%%),'
Ã#"1$%$

Figure 1: PRISMA flow diagram for the selection of systematic reviews. *From:* Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: http://www.prisma-statement.org/

Outcomes:

1. **Puberty blockers:** We found 4 systematic reviews assessing the effects of puberty blockers published between 2020 and 2022.[1-4] From these, we judged 2 as having moderate methodological quality, and 2 as having critically low methodological quality. Details of the assessment are provided in Figure 2.

   Table 1 summarizes the evidence about the effects of puberty blockers on the outcomes of interest. We used information from 2 systematic reviews.[2 3] For most outcomes (except suicidality), there is no evidence about the effect of puberty blockers compared to not using puberty blockers. In other words, no studies compared the outcomes between a group of people with gender dysphoria using puberty blockers and another not using them. Therefore, it is unknown whether people with gender dysphoria who use puberty blockers experience more improvement in gender dysphoria, depression, anxiety, and quality of life than those with gender dysphoria who do not use them. There is very low certainty about the effects of puberty blockers on suicidal ideation (see details in Table 1).

   Studies, however, reported outcomes among a group of people with gender dysphoria after receiving puberty blockers. The findings are:

   - There is low certainty evidence suggesting that treatment with puberty hormones may slightly increase gender dysphoria severity (mean change score in the Utrecht Gender Dysphoria scale, 0.7 points [95% CI, -4.2 to 5.6], range 12-60, with higher scores reflecting more severe gender dysphoria)
   - There is low certainty evidence suggesting that treatment with puberty blockers may decrease depression (mean change score in the Beck Depression Inventory, -3.4 [95% CI, -5.7 to -1.0], range 0-63, with higher scores reflecting more severe depression)
   - There is low certainty evidence suggesting that treatment with puberty blockers may decrease anxiety (mean change score in the Trait Anxiety Scale, trait subscale, -1.5 [95% CI, -4.7 to -1.8], range 0-80, with higher scores reflecting more severe anxiety)
   - There is low certainty evidence suggesting a moderate percentage of patients reporting adverse events after treatment with puberty blockers (see Table 1 for details)
   - There is very low certainty evidence about how puberty blockers affect suicidality

Figure 2: AMSTAR assessment judgements for systematic reviews addressing puberty blockers

Review ID	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Methodological quality
AHRQ 2021																	MODERATE
NICE 2020a																	MODERATE
Ramos 2020																	CRITICALLY LOW
Rew 2020																	CRITICALLY LOW

Figure legend:

Yes

Probably yes

Probably no

No

Not applicable

Table 1: Puberty blockers (gonadotrophin releasing hormone analogues) compared to no puberty blockers in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: puberty blockers (gonadotrophin releasing hormone analogues)
Comparison: no puberty blockers

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no puberty blockers	Risk / mean with puberty blockers				
Gender dysphoria assessed with: difference (effect) in gender dysphoria proportion or severity	Not reported					The effects of puberty blockers on gender dysphoria are unknown
Gender dysphoria assessed with: mean change score in the Utrecht Gender Dysphoria Scale (12-60, higher scores reflect more gender dysphoria, 40 points or more indicate a diagnosis of gender dysphoria) (NICE, 2020a) Follow up: mean 1.9 years (range 0.4 to 5.1 years)	NA	0.7 (-4.2 to 5.6)	NA	41 (1 study)	⨁⨁◯◯ LOW[1]	The mean gender dysphoria score may increase by 0.7 points after puberty blockers
Depression assessed with: difference (effect) in depression proportion or severity	Not reported					The effects of puberty blockers on depression are unknown

App. 089

Table 1: Puberty blockers (gonadotrophin releasing hormone analogues) compared to no puberty blockers in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: puberty blockers (gonadotrophin releasing hormone analogues)
Comparison: no puberty blockers

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no puberty blockers	Risk / mean with puberty blockers				
Depression assessed with: mean change score in Beck Depression Inventory-II scale (0-63, higher scores represent more severe depression) (NICE, 2020a) Follow up: mean 1.9 years (range 0.4 to 5.1 years)	NA	-3.4 (-5.7 to -1.0)	NA	41 (1 study)	⨁⨁◯◯ LOW[1]	The mean depression score may decrease by 3.4 points after puberty blockers
Anxiety assessed with: difference (effect) in anxiety proportion or severity	Not reported					The effects of puberty blockers on anxiety are unknown
Anxiety assessed with: mean change score in STAI-Trait scale (0-80, higher scores represent more severe anxiety) (NICE, 2020a) Follow up: mean 1.9 years (range 0.4 to 5.1 years)	NA	-1.5 (-4.7 to 1.8)	NA	41 (1 study)	⨁⨁◯◯ LOW[1]	The mean anxiety score may decrease by 1.5 points after puberty blockers

App. 090

Table 1: Puberty blockers (gonadotrophin releasing hormone analogues) compared to no puberty blockers in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: puberty blockers (gonadotrophin releasing hormone analogues)
Comparison: no puberty blockers

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no puberty blockers	Risk / mean with puberty blockers				
Quality of life assessed with: any measure	Not reported					The effects of puberty blockers on quality of life are unknown
Suicidal ideation difference (effect) in suicidal ideation (Rew, 2020) Follow-up: cross-sectional survey	The authors report that "compared to youth who did not receive pubertal suppression, those who did showed lower lifetime rates of suicidal ideation".			89 (1 study)	⊕◯◯◯ VERY LOW[2]	We are very uncertain about the effect of puberty blockers on suicidal ideation
Adverse effects assessed with: proportion of patients reporting adverse effects (NICE, 2020a) Follow up: mean 2.3 years (range 0.0 to 11.3 years)	NA	11%[3] (2% to 29%)	NA	27 (1 study)	⊕⊕◯◯ LOW[4]	The proportion of patients reporting adverse effects after treatment with puberty blockers may be 11%

STAI-Trait: Trait Anxiety Scale. Range: 0-80
CI: Confidence interval
NA: Not applicable

Table 1: Puberty blockers (gonadotrophin releasing hormone analogues) compared to no puberty blockers in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: puberty blockers (gonadotrophin releasing hormone analogues)
Comparison: no puberty blockers

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no puberty blockers	Risk / mean with puberty blockers				

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

1. Ā Mean change rated down due to risk of bias and imprecision. According to the systematic review authors, the study had poor methodological quality. In addition, there are too few participants included, which is not sufficient to make trustworthy inferences (does not meet the optimal information size).
2. Ā The authors of Rew 2020 narratively summarized the outcome of Turban *et al.* 2020; a cross-sectional online survey study. According to the systematic review authors, Turban et al. did not describe the study participants and the setting in detail and it was unclear whether outcomes were measured in a valid and reliable way. We therefore, downgraded the certainty of evidence by one level from low to very low due to high risk of bias.
3. Ā The authors reported 3/27 (11%) participants treated with GnRHa developed side effects: 1 participant developed sterile abscesses; they were switched from leuprolide acetate to triptorelin, 1 participant developed leg pains and headaches, which eventually resolved without treatment, 1 participant gained 19 kg within 9 months of initiating GnRH analogues.
4. Ā Proportion of adverse effects rated down due to risk of bias and imprecision. According to the systematic review authors, the cohort study Khatchadourian *et al.* 2014 was assessed at high risk of bias due to incomplete reporting of its cohort. In addition, there are too few participants included, which is not sufficient to make trustworthy inferences (does not meet the optimal information size).

2.ĀCross-sex hormones: We found 9 systematic reviews assessing the effects of cross-sex hormones published between 2020 and 2022.[4-12] One of these, however, included both puberty blockers and cross-sex hormones combined in their evidence synthesis as was not prioritized.[5] From the 8 remaining reviews, we judged 1 as having high methodological quality, 2 as having moderate methodological quality, 2 as having low methodological quality, and 3 as having critically low methodological quality. Details of the assessment are provided in Figure 3. Because of its eligibility criteria related to study design, the systematic review judged at high methodological quality[7] did not include any studies and therefore we could not use it to inform any outcome.

Table 2 summarizes the evidence about the effects of cross-sex hormones on the outcomes of interest. We used information from 4 systematic reviews.[6 9 11 12] For most outcomes (all except risk of breast cancer), there is no evidence about the effect of cross-sex hormones compared to not using cross-sex hormones. In other words, no studies compared the outcomes between a group of people with gender dysphoria using cross-sex hormones and another not using it. Therefore, it is unknown whether people with gender dysphoria who use cross-sex hormones experience more improvement in gender dysphoria, depression, anxiety, quality of life, and suicidality than those with gender dysphoria who do not use them. There is low certainty evidence suggesting that cross-sex hormones may not increase or decrease the risk of breast cancer (see details in Table 2).

Studies, however, reported outcomes among a group of people with gender dysphoria after receiving cross-sex hormones. The findings are:

-Ā There is low certainty evidence suggesting that treatment with cross-sex hormones may decrease gender dysphoria severity (mean change score in the Utrecht Gender Dysphoria scale, -42.4 points [95% CI, -44.1 to -40.1], range 12-60, with higher scores reflecting more severe gender dysphoria)
-Ā There is low certainty evidence suggesting that treatment with cross-sex hormones may decrease depression (measured with different scales, see Table 4 for details) and the need for treatment for depression (change in percentage, -39%)
-Ā There is low certainty evidence suggesting that treatment with cross-sex hormones may decrease anxiety (measured with different scales, see Table 4 for details) and the need for treatment for anxiety (change in percentage, -32%)
-Ā There is very low certainty about the change in quality of life after treatment with cross-sex hormones.
-Ā There is low certainty evidence suggesting that treatment with cross-sex hormones may decrease suicidality degree (mean change score in the Ask Suicide-Screening questions scale, -0.84 points [95% CI, -1.30 to -0.44], range 0-4, with higher scores reflecting more severe suicidality) and the percentage of patients with need for treatment due to suicidality/self-harm (change in percentage, -31%). There is very low certainty evidence about the percentage of people with suicidal ideation and suicide attempts after treatment with cross-sex hormones.

- There is low certainty evidence suggesting a low prevalence of venous thromboembolism after treatment with cross-sex hormones (see Table 2 for details)

Figure 3: AMSTAR assessment judgements for systematic reviews addressing cross-sex hormones

Review ID	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Methodological quality
AHRQ 2021																	MODERATE
Baker 2021																	MODERATE
Fledderus 2020																	CRITICALLY LOW
Haupt 2020																	HIGH
Karalexi 2020																	LOW
Kotamarti 2021																	CRITICALLY LOW
Mattawanon 2021																	CRITICALLY LOW
NICE 2021b																	MODERATE
Totaro 2021																	LOW

Figure legend:

Yes

Probably yes

Probably no

No

Not applicable

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				
Gender dysphoria assessed with: difference (effect) in gender dysphoria percentage or severity	Not reported					The effects of cross-sex hormones on gender dysphoria are unknown
Gender dysphoria assessed with: mean change score in the Utrecht Gender Dysphoria Scale (12-60, higher scores reflect more gender dysphoria, 40 points or more indicate a diagnosis of gender dysphoria) (NICE, 2020b) Follow up: 1 year	NA	-42.4 (-44.1 to -40.1)	NA	23 (1 study)	⨁⨁◯◯ LOW[1]	The mean gender dysphoria score may decrease by 42 points after cross-sex hormones
Depression assessed with: difference (effect) in depression percentage or severity	Not reported					The effects of cross-sex hormones on depression are unknown

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				
Depression assessed with: mean change score in depression scales (higher scores represent more severe depression) (NICE, 2020b) Follow up: 1 year	NA	The mean depression score reduction was 9.6 points when using the BDI-II scale (n=23) and 7.5 when using the CESD-R scale (n=50). The authors report that both reductions were statistically significant[2]	NA	73 (2 studies)	⊕⊕◯◯ LOW[1]	The mean depression score may decrease after cross-sex hormones
Depression assessed with: change in percentage of patients with need for treatment (NICE, 2020b) Follow-up: 1 year	NA	The percentage of participants requiring treatment was reduced by 39% (from 54% at baseline), which was statistically significant	NA	52 (1 study)	⊕⊕◯◯ LOW[1]	The percentage of participants requiring treatment may be reduced by 39% after cross-sex hormones
Anxiety assessed with: difference (effect) in anxiety percentage or severity	Not reported					The effects of cross-sex hormones on anxiety are unknown

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				
Anxiety assessed with: mean change score in anxiety scales (higher scores represent more severe anxiety) (NICE, 2020b) Follow up: 1 year	NA	The mean anxiety score reduction was 16.5 points when using the STAI-State scale and 14.5 when using the STAI-Trait scale. The authors report that both reductions were statistically significant	NA	23 (1 study)	⊕⊕◯◯ LOW[1]	The mean anxiety score may decrease after cross-sex hormones
Anxiety assessed with: change in percentage of patients with need for treatment (NICE, 2020b) Follow-up: 1 year	NA	The percentage of participants requiring treatment was reduced by 32% (from 48% at baseline), which was statistically significant	NA	52 (1 study)	⊕⊕◯◯ LOW[1]	The percentage of participants requiring treatment may be reduced by 32% after cross-sex hormones
Quality of life assessed with: difference (effect) in quality of life improvement	Not reported					The effects of cross-sex hormones on quality of life are unknown

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				
Quality of life assessed with: mean change score in QLES-Q-SF score (higher scores represent better quality of life) (NICE, 2020b) Follow up: 1 year	NA	The mean quality of life score improved, but the differences were not statistically significant. The magnitudes were not reported	NA	50 (1 study)	⨁◯◯◯ VERY LOW[3]	We are very uncertain about the quality of life change after cross-sex hormones
Suicide/ suicidal ideation assessed with: difference (effect) in suicide or suicidal ideation	Not reported					The effects of cross-sex hormones on suicide/ suicidal ideation are unknown
Suicidality assessed with: change in score from ASQ instrument (higher scores represent greater degree of suicidality) (NICE, 2020b) Mean follow up: 1 year	NA	-0.84 (-1.30 to -0.44)	NA	39 (1 study)	⨁⨁◯◯ LOW[1]	Suicidality scores may decrease by 0.84 points after cross-sex hormones

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				
Suicidal ideation assessed with: percentage of participants with suicidal ideation measured with PHQ-9 (NICE, 2020b) Follow-up: 1 year	NA	The percentage of participants with suicidal ideation decreased by 6% (from 10% at baseline). The authors did not conduct a statistical analysis	NA	50 (1 study)	⨁◯◯◯ VERY LOW[3]	We are very uncertain about the change in percentage of patients in suicidal ideation after cross-sex hormones
Suicide attempts assessed with: not reported (NICE, 2020b) Follow up: not reported	NA	The percentage of people with lifetime suicide attempts was 15%, those with attempts 3 months before treatment was 2%, and those with attempts at follow up was 5%	NA	130 (1 study)	⨁◯◯◯ VERY LOW[3]	We are very uncertain about the percentage of people with suicide attempts after cross-sex hormones
Suicidality/ self-harm assessed with: change in percentage of patients with need for treatment (NICE, 2020b) Follow-up: 1 year	NA	The percentage of participants requiring treatment was reduced by 31% (from 35% at baseline), which was statistically significant	NA	52 (1 study)	⨁⨁◯◯ LOW[1]	The percentage of participants requiring treatment may be reduced by 31% after cross-sex hormones
Venous thromboembolism assessed with: Risk of VTE	Not reported					The effects of cross-sex hormones on the risk of VTE are unknown

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				
Venous thromboembolism assessed with: Prevalence among assigned males at birth (Totaro, 2021) Mean follow up: 4.1 years	NA	**20 per 1,000** (10 to 30)	NA	11,542 (18 studies)	⊕⊕⊕◯ MODERATE[4]	The prevalence of VTE among assigned males at birth is probably 2% after cross-sex hormones
Venous thromboembolism assessed with: Prevalence among assigned females at birth (Kotamarti, 2021) Mean follow up: 5.7 years	NA	**6 per 1,000** (CI not reported)[5]	NA	4,218 (8 studies)	⊕⊕⊕◯ MODERATE[6]	The prevalence of VTE among assigned females at birth is probably 0.6% after cross-sex hormones
Breast cancer assessed with: Risk of breast cancer (Fledderus, 2020) Follow up: not reported	Two studies compare the risk of beast cancer between assigned females at birth using versus not using testosterone, and found no differences (0 vs 1 case [total n= 130], and 1 vs 6 [total n=1579]). A third study compared assigned females at birth with non transgender women and found a lower risk in the former (magnitude not reported)		NA	2,938 (3 studies)	⊕⊕◯◯ LOW[7]	The risk of breast cancer may not increase or decrease due to the use of cross-sex hormones

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				

ASQ: Ask Suicide-Screening Questions. Range 0-4
BDI-II: Beck Depression Inventory. Range: 0-63
CESD-R: Center for Epidemiological Studies Depression Scale. Range: 0-60
CI: Confidence interval
NA: Not applicable
PHQ-9: Patient Health Questionnaire (PHQ) Modified for Teens. For suicidal ideation, it is a single question (yes/no)
QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire. Range: 15-75
STAI: State-Trait Anxiety Inventory. Range: 0-80

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

1. Ā Mean change rated down due to risk of bias and imprecision. According to the systematic review authors, the studies had poor methodological quality. In addition, there are too few participants included, which is not sufficient to make trustworthy inferences (does not meet the optimal information size)
2. Ā Similar results when this outcome was measured using the Patient Health Questionnaire (PHQ) Modified for Teens in one of the same studies
3. Ā Rated down due to risk of bias, imprecision, and indirectness. According to the systematic review authors, the studies had poor methodological quality. In addition, there are too few participants included, which is not sufficient to make trustworthy inferences (does not meet the optimal information size). Finally, 30% of the participants did not have a diagnosis of gender dysphoria.
4. Ā Prevalence rated down due to risk of bias. According to the systematic review authors, only 6 out of the 18 studies (representing 16.5% of the weight of the studies) were at low risk of bias.

5.Ā A meta-analysis of independent studies reported in this systematic review suggested that the prevalence of VTE in non-transgender females at birth was 1.7% (based on 7 studies and 18,748 persons)
6.Ā Prevalence rated down doe to risk of bias. According to the systematic review authors, all studies had at least one domain judged as problematic.
7.Ā Risk rated down 2 levels because of risk of bias. The researchers did not account for confounding in any of the studies.

3.ĀSurgeries: We found 15 systematic reviews assessing the effects of gender-affirming surgeries published between 2020 and 2022. We judged 8 as having low methodological quality and 7 as having critically low methodological quality. Details of the assessment are provided in Figure 4. We present the results regarding the effects of surgeries in three parts. First, we describe the effects of all surgeries on mental health outcomes in all patients. Second, we describe the effects of all surgeries on surgical outcomes in assigned females at birth (transgender males). Finally, we describe the effects of all surgeries on surgical outcomes in assigned males at birth (transgender females).

3.1ĀEffects of surgeries on mental health outcomes: Table 3 summarizes the evidence about the effects of all surgeries on mental health outcomes in all patients. We used information from 2 systematic reviews.[13 14] There were no systematic reviews and studies reporting on gender dysphoria, depression, anxiety, and suicidality. Therefore, the effects of surgeries on these outcomes (when compared to a group of patients with gender dysphoria who do not undergo surgery), or the changes in these outcomes (improvements or deterioration) among patients who undergo surgeries is unknown.

The systematic reviews addressed quality of life and depression, but none of the included studies included a comparison group. Thus, it is unknown whether people with gender dysphoria who undergo surgeries experience more improvement in quality of life or less regret than those with gender dysphoria who do not undergo surgeries.

Studies, however, reported the following outcomes among a group of people with gender dysphoria after undergoing surgeries. The findings are:

-Ā There is low certainty evidence suggesting that the percentage of people who experience regret after surgery is low (1%)
-Ā There is very low certainty evidence about how surgeries affect quality of life (see Table 3 for details)

Figure 4: AMSTAR assessment judgements for systematic reviews addressing gender-affirming surgery

Review ID	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Methodological quality
Bustos SS 2021																	LOW
Bustos VP 2021																	LOW
Bustos VP 2021b																	LOW
Dunford 2021																	LOW
Eftekhar, 2020																	LOW
Falcone 2021																	CRITICALLY LOW
Hu, 2022																	CRITICALLY LOW
Huayllani 2021																	CRITICALLY LOW
Jolly 2021																	LOW
Nassiri 2020																	CRITICALLY LOW
Oles 2022																	LOW
Oles 2022b																	LOW
Salibian 2021																	CRITICALLY LOW
Sijben 2021																	CRITICALLY LOW
Tay 2021																	CRITICALLY LOW

Figure legend:

Yes

Probably yes

Probably no

No

Not applicable

Table 3: All surgeries compared to no surgeries in young people (<21 years old) with gender dysphoria

Patient or population: young people (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries
Outcomes: Mental health and regret

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Gender dysphoria assessed with: any measure	Not reported					The effects of surgery on gender dysphoria, the changes in gender dysphoria severity after surgery, and the prevalence of gender dysphoria after surgery are unknown
Depression assessed with: any measure	Not reported					The effects of surgery on depression, the changes in depression severity after surgery, and the prevalence of depression after surgery are unknown
Anxiety assessed with: any measure	Not reported					The effects of surgery on anxiety, the changes in anxiety severity after surgery, and the prevalence of anxiety after surgery are unknown
Suicidality assessed with: any measure	Not reported					The effects of surgery on suicidality, the changes in anxiety severity after surgery, and the prevalence of anxiety after surgery are unknown
Quality of life assessed with: difference (effect) in quality of life	Not reported					The effects of surgery on quality of life are unknown
Quality of life assessed with: change in quality of life	Not reported					The change in quality of life after surgery is unknown

Quality of life assessed with: mean score in the Short Form-36 Scale (0-100, higher scores reflect better quality of life) (Eftekhar Ardebili, 2020) Follow up: cross-sectional	NA	59.17 (48.59 to 69.74)[1]	NA	633 (5 studies)	⊕◯◯◯ VERY LOW[2]	We are very uncertain about the quality of life after surgeries
Regret assessed with: difference (effect) in percentage of people with regret	Not reported					The effects of surgery on regret are unknown
Regret assessed with: percentage of people with regret (Bustos, 2021) Mean follow up: 4 years	NA	1% (0 to 2%)[3]	NA	7928 (27 studies)	⊕⊕◯◯ LOW[4]	The percentage of people who experience regret is low

CI: Confidence interval
NA: Not applicable

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

1. Ā Similar scores for assigned males at birth and assigned females at birth.
2. Ā Mean score rated down for risk of bias and inconsistency. According to the systematic review authors, all studies had concerns related to risk of bias. In addition, the smaller studies showed better quality of life than the larger study.
3. Ā Similar percentage for assigned males at birth and assigned females at birth, and for different types of surgeries (all pooled percentages below 2%).
4. Ā Percentage rated down due to risk of bias and indirectness. According to the authors, many of the studies had moderate or high risk of bias. The mean age of the participants at the time of surgery was higher than the target population. Because it was considered to not have an important effect on the pooled estimate, we did not rate down for statistical heterogeneity

3.2 **Effects of surgeries on assigned females at birth:** Table 4 summarizes the evidence about the effects of all surgeries on surgical outcomes among assigned at birth females. We used information from 3 systematic reviews.[13-17] Due to the nature of the outcomes (i.e. they can only be experienced by people who undergo surgeries), there cannot be studies comparing the outcomes between a group of people with gender dysphoria who undergo surgeries and another who does not.

Studies, therefore, assessed the outcomes among a group of people with gender dysphoria after surgery. The findings are:

-Ā There is low certainty evidence suggesting that the percentage of people who are satisfied after chest surgery is high (92%)
-Ā There is very low certainty evidence about the rate of surgical complications after chest surgery
-Ā There is very low certainty evidence about the percentage of people who are satisfied, and the rate of surgical complications after bottom surgeries (see Table 4 for details)

Table 4: All surgeries compared to no surgeries in assigned females at birth (<21 years old) with gender dysphoria

Patient or population: assigned females at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI) Risk / mean with no surgery	Risk/ mean with surgery	Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
Chest surgery						
Satisfaction assessed with: percentage of people who reported being satisfied (Bustos VP, 2020b) Range of follow up: 6 weeks to 46 months[1]	NA	92% (88% to 96%)[2]	NA	733 (14 studies)	⨁⨁◯◯ LOW[3]	The percentage of people who reports being satisfied may be 92%
Surgical complications assessed with: rate of complications across patients (Oles, 2022) Range of follow up: 8 weeks to 1 year	NA	16.8% Range (5.5% to 80.0%)	NA	1255 (7 studies)	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the rate of surgical complications
Reoperation assessed with: rate of reoperation across patients (Oles, 2022) Range of follow up: 8 weeks to 1 year	NA	6.2% Range (0.7% to 11.2%)	NA	1214 (6 studies)	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the rate of reoperation
Bottom surgery						

Table 4: All surgeries compared to no surgeries in assigned females at birth (<21 years old) with gender dysphoria

Patient or population: assigned females at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Satisfaction assessed with: percentage of people who reported being satisfied (Oles, 2022b) Range of follow up: 6 weeks to 46 months	NA	89.6% (45% to 100%)[5]	NA	1458 (27 studies)	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the percentage of people who reports being satisfied
Surgical complications-Major assessed with: percentage of people experiencing major complications (Oles, 2022b) follow up: not reported	NA	The percentage was - 2.3% (range 0 to 20%) experiencing total flap loss - 19.5% (range 0 to 72%) experiencing prosthesis issues - 24.5% (range 0 to 86%) experiencing urethral issues	NA	3177 (42 studies)[6]	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the percentage of people who experience major surgical complications
Surgical complications-Minor assessed with: percentage of people experiencing major complications (Oles, 2022b) follow up: not reported	NA	The percentage varied from 9.3% (range 0% to 45.5%) experiencing donor site issues, to 24% (range 10 to 93%) experiencing urethral issues[7]	NA	4466 (52 studies)[8]	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the percentage of people who experience minor surgical complications

Table 4: All surgeries compared to no surgeries in assigned females at birth (<21 years old) with gender dysphoria

Patient or population: assigned females at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Reoperation assessed with: rate of reoperation across patients (Oles, 2022b) follow up: not reported	NA	27.6% Range (2.5% to 40%)	NA	1624 (15 studies)	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the percentage of people who undergo reoperations

CI: Confidence interval
NA: Not applicable

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

1. Ā Studies used different scales to assess satisfaction
2. Ā The percentage was similar when the analysis was done by type of surgery and by follow up time (< 1 year vs 1 year or more). Another systematic review (Oles, 2022) also investigated this outcome, and reported a very similar percentage of satisfaction (91.8%, range 73% to 100%)
3. Ā Percentage of patients satisfied rated down due to risk of bias and indirectness. According to the systematic review authors, several studies were judged at moderate and high risk of bias. In addition, the median of the mean age of patients included in the studies was 28 years
4. Ā Rated down due to risk of bias, inconsistency/ imprecision, and indirectness. Even though the review authors did not assess risk of bias, these studies were included in other systematic reviews in which the authors judged several of them at high risk of bias. The studies report inconsistent results (some high and other low rates). The patients are older than the target population.
5. Ā Results for phalloplasty. Similar results for metoidioplasty (91.3%).

6.Ā People and studies for urethral complications. 2671 people (37 studies) for prosthesis issues, and 1548 people (22 studies) for total flap loss.
7.Ā Percentage of wound dehiscence 9.8% (range, 2.9% to 75%), percentage of infection/ partial necrosis 10.3% (range, 0 to 45.8%), percentage of prosthesis issues 14.2% (range, 1.6 to 41.9%), percentage of incontinence 15.3% (range, 5.4% to 59.1%)
8.Ā People and studies for infection/ partial necrosis. 2389 people (31 studies) for urethral issues, 1736 people (17 studies) for wound dehiscence, 1080 (10 studies) for prosthesis issues, 1053 people (8 studies) for donor site issues, 131 people (3 studies) for incontinence

3.3 **Effects of surgeries on assigned males at birth**: Table 5 summarizes the evidence about the effects of all surgeries on surgical outcomes among assigned at birth males. We used information from 3 systematic reviews.[16 18 19] Due to the nature of the outcomes (i.e. they can only be experienced by people who undergo surgeries), there cannot be studies comparing the outcomes between a group of people with gender dysphoria who undergo surgeries and another who does not.

Studies, therefore, assessed the outcomes among a group of people with gender dysphoria after surgery. The findings are:

-Ā There is low certainty evidence suggesting that the percentage of people who are satisfied after vaginoplasty is high (91%)
-Ā There is very low certainty evidence about the percentage of people who are satisfied, the rate of complications, and the rate of reoperations after chest surgery (see Table 5 for details)
-Ā There is low certainty evidence suggesting that the percentage of people who have regret after vaginoplasty is low (2%)
-Ā There is very low certainty evidence about the rate of complications and the rate of reoperations after vaginoplasty (see Table 5 for details)

Table 5: All surgeries compared to no surgeries in assigned males at birth (<21 years old) with gender dysphoria

Patient or population: assigned males at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Chest surgery						
Satisfaction assessed with: percentage of people who reported being satisfied (Oles 2022) Range of follow up: 12 months to 17 years	NA	Range 75% (80/107) to 95% (33/35)[1]	NA	142 (2 studies)	⨁◯◯◯ VERY LOW[2]	We are very uncertain about the percentage of people who report being satisfied
Surgical complications assessed with: rate of complications across patients (Oles 2022) Range of follow up: 2 weeks to 16 years	NA	The complication rates were: - 3.8% (range 0% to 5.5%) of capsular contracture - 2.2% of major hematoma - 2.2% of implant extrusion[3]	NA	432 (5 studies)	⨁◯◯◯ VERY LOW[2]	We are very uncertain about the rate of surgical complications
Reoperation assessed with: rate of reoperation across patients (Oles 2022) Range of follow up: Not reported	NA	8.6% Range (4.4% to 10.4%)	NA	291 (2 studies)	⨁◯◯◯ VERY LOW[2]	We are very uncertain about the rate of reoperation
Bottom surgery						

Table 5: All surgeries compared to no surgeries in assigned males at birth (<21 years old) with gender dysphoria

Patient or population: assigned males at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Satisfaction assessed with: percentage of people who reported being satisfied for overall outcomes (Bustos SS, 2021) Range of follow up: 1 week to 11.3 years	NA	91% (81% to 98%)[4]	NA	1230 (12 studies)	⨁⨁◯◯ LOW[5]	The percentage of people who report being satisfied with overall outcomes may be 91%
Regret assessed with: percentage of people who reported regret (Bustos SS, 2021) Range of follow up: 2 months to 24.1 years	NA	2% (1% to 3%)	NA	1137 (15 studies)	⨁⨁◯◯ LOW[6]	The percentage of people who report regret may be 2%
Surgical complications assessed with: rate of complications across patients (Bustos SS, 2021) Range of follow up: 3 weeks to 24.1 years	NA	The complication rates were: - 1% (95% CI, <0.1% to 2%) of fistula - 11% (95% CI, 8% to 14%) of stenosis and/or strictures - 4% (95% CI, 1% to 9%) of tissue necrosis - 3% (95% CI, 1% to 4%) of prolapse[7]	NA	4196 (42 studies)[3]	⨁◯◯◯ VERY LOW[8]	We are very uncertain about the rate of surgical complications

Table 5: All surgeries compared to no surgeries in assigned males at birth (<21 years old) with gender dysphoria

Patient or population: assigned males at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Reoperation assessed with: rate of reoperation across patients (Tay, 2021) Range of follow up: 6 weeks to 14.8 months	NA	One study reported a surgical revision rate of 9% (1/11 patients), and a second study reported that 13% (19/145) patients required repeat surgery due to complications.	NA	156 (2 studies)	⨁◯◯◯ VERY LOW[9]	We are very uncertain about the percentage of people who undergo reoperations

CI: Confidence interval
NA: Not applicable

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

1.Ã Another systematic review, Sijben 2021, reported satisfaction from 3 additional studies: 82% (113/138) were satisfied or very satisfied, 93% (32/34) were happier and more satisfied with their chest, and 79% (28/36) were very satisfied with the overall cosmetic result (very low certainty of evidence due to risk of bias, imprecision, and indirectness).
2.Ã Rated down due to risk of bias, indirectness (the included studies were not restricted to youth or young adults), and imprecision (too few participants included, not meeting optimal information size).

3.Ā Another systematic review, Sijben 2021, reported similar ranges for rates of complication requiring reoperation from 7 studies (835 patients): capsular contraction (range 0.0-5.6%), asymmetry (3.6%), hematoma (range 0.0-2.9%), infection (range 0.0-0.9%), striae distensae (0.7%), implant rupture (0.7%), abscess (0.4%), scarring (0.0%), hypersensitivity (0.0%), and numbness (0.0%) (very low certainty of evidence due to risk of bias, imprecision, and indirectness)
4.Ā Bustos SS *et al.* 2021 additionally reported on satisfaction for functional (87%, 95% CI 77% to 94%) and aesthetic (90%, 95% CI 84% to 94%) outcomes. Another systematic review and meta-analysis, Oles 2022b, similarly reported that 92.3% (range 23.1% to 100%) of patients (2410/2601) were satisfied after vaginoplasty (very low certainty of evidence due to risk of bias, imprecision, and indirectness).
5.Ā Rated down due to risk of bias (the systematic review authors reported the quality of the included studies to be low to moderate using the New Castle Ottawa scale), and indirectness as the included studies were not restricted to youth or young adults. We did not rate down for imprecision or inconsistency despite high I^2 values as a satisfaction rate of 80% or above was deemed as a minimum threshold for clinical importance.
6.Ā Rated down due to risk of bias (the systematic review authors reported the quality of the included studies to be low to moderate using the New Castle Ottawa scale), and indirectness as the included studies were not restricted to youth or young adults.
7.Ā Another systematic review, Oles 2022b, similarly reported the percentage of patients experiencing complications from 51 studies, ranging from 2.4% to 12.0% (range 0% to 88%) for minor complications (intraoperative injury, wound dehiscence, superficial necrosis, infection, urinary issues, vaginal prolapse, stenosis, and bleeding) and 1.6% to 2.1% (range 0% to 31%) for major complications (flap/graft necrosis and infection) after genitoplasty (very low certainty of evidence due to risk of bias, imprecision, and indirectness).
8.Ā Rated down due to risk of bias (the systematic review authors reported the quality of the included studies to be low to moderate using the New Castle Ottawa scale), imprecision and inconsistency, with wide confidence intervals and I^2 values ranging from 65.8% to 94.3%, and indirectness as the included studies were not restricted to youth or young adults.
9.Ā Rated down due to risk of bias, indirectness (the age range of patients in the included studies was 24 to 39 years; the studies included were restricted to those that investigated the use of peritoneum in neovagina construction), and imprecision (too few participants included, not meeting optimal information size).

Results from search for studies not included in the systematic reviews: After screening 1854 records found through our searches, we found 10 eligible studies (figure 5). From these, 8 were comparative observational studies[20-27] and 2 were non- comparative[28 29]. We provide reasons for excluding studies in appendix 2.

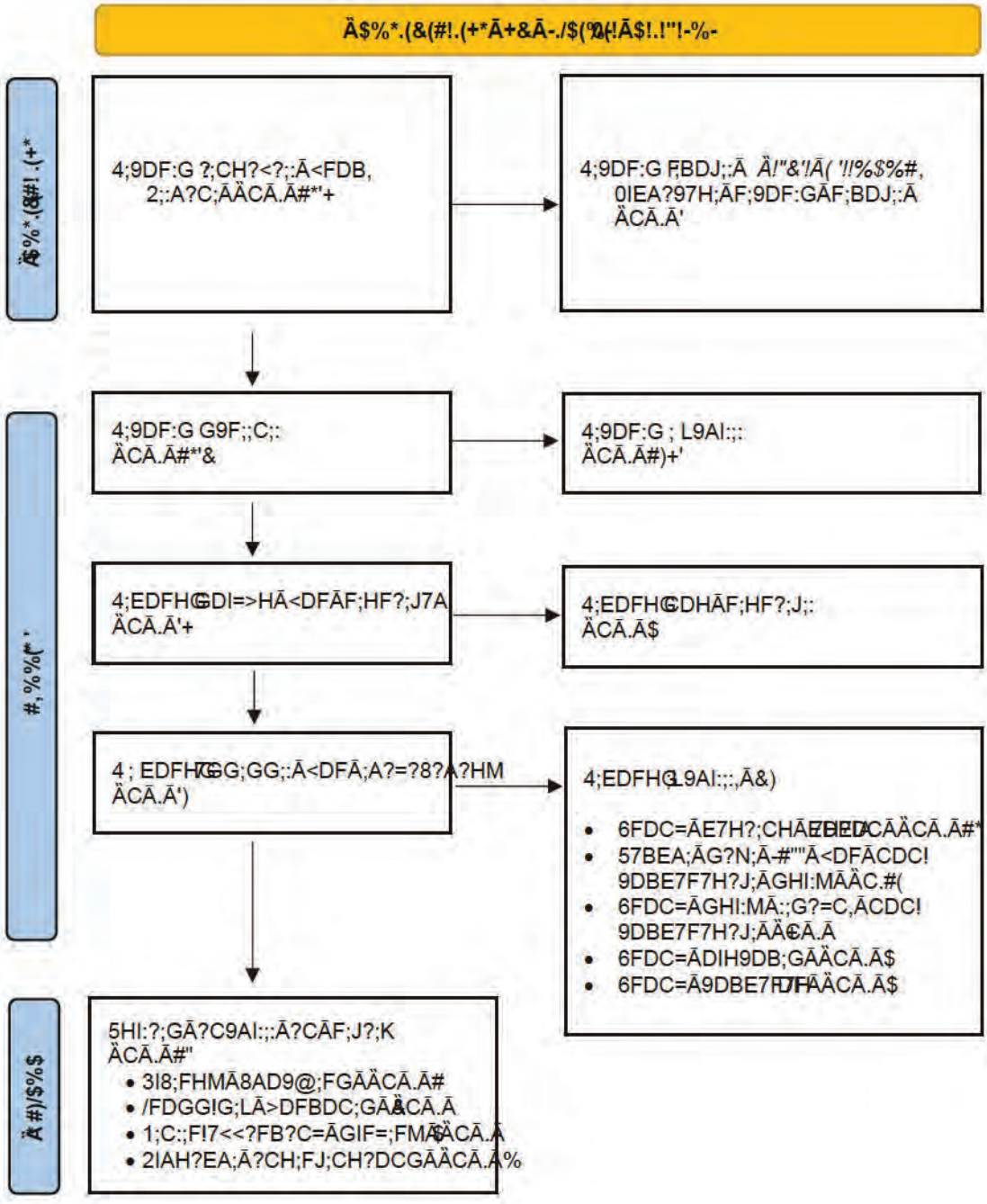

Figure 5: PRISMA flow diagram for the selection of primary studies. *From:* Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: http://www.prisma-statement.org/

None of the studies were judged as likely to importantly change the conclusions obtained from the systematic reviews (Tables 6 and 7). The main limitations of the comparative studies were risk of bias concerns (Figures 6 and 7) due to confounding, classification of intervention, and missing data; as well as small sample sizes. Although non-comparative studies were at lower risk of bias, because their results were consistent with those of the included evidence, they were also judged as unlikely to change the conclusions importantly.

Table 6: Characteristics of eligible comparative observational studies

Study ID	Sample size*	Study design	Intervention	Comparator	Outcomes measured	Likely to change conclusions	Reasons
VanDerMiesen, 2020	450	Retrospective cohort study	Puberty blockers	Waiting for puberty blockers	Self-harm/ suicidality, internalizing behaviors	No	Reports a small benefit on suicidality and moderate on internalizing behaviours, but high risk of bias
Becker-Hebly, 2021	75	Prospective cohort study	1. Puberty blockers 2. Cross-sex hormones 3. Surgery	No medical intervention yet; psychosocial intervention only	Health-related quality of life	No	Critical risk of bias (missing data due to low response rate, and confounding). Reports small benefit in mean change score for mental and physical dimension QoL as compared to no medical treatment. Imprecision; the 95% CIs for mean change scores are wide.
Green, 2021	3235	Cross-sectional study	Cross-sex hormones	Would like to take cross-sex hormones	Depression, suicidality	No	Critical risk of bias, no follow up of patients (measurement of current outcomes and not adjusting for baseline)
Tordoff, 2022	84	Prospective cohort study	1. Puberty blockers 2. Cross-sex hormones	No intervention	Depression, anxiety, suicidal thoughts	No	Moderate risk of bias, small sample size
Turban, 2022	9341	Cross-sectional study	Cross-sex hormones	Desired but never accessed gender affirming hormones	Suicidal ideation, suicidal attempt	No	Critical risk of bias, no follow up of patients (measurement of current outcomes and not adjusting for baseline)
Grannis, 2021	47	Cross-sectional study	Cross-sex hormones	No intervention yet	Anxiety, depression	No	Critical risk of bias, no follow up of patients, small sample size
Fontanari, 2020	350	Cross-sectional study	1. Cross-sex hormones 2. Cross-sex hormones or surgery	1. Waiting for cross-sex hormones 2. No intervention	Anxiety, depression, gender distress	No	Critical risk of bias (confounding, self-reported classification of interventions). Online cross-sectional survery reported small benefit in anxiety and depression mean scores, and little to no effect on gender distress with cross-sex hormones and/or surgery. Non-randomized comparative study provides very low certainty evidence due to

							very serious risk of bias and serious imprecision (95% CIs include little to no effect)
Castelo-Branco, 2021	205	Cross-sectional study	Cross-sex hormones	No intervention	Anxiety, depression	No	Critical risk of bias due to confounding (non-adjusted analysis). Reported no difference observed in anxiety and depression mean scores (Symptom Checklist-90-Revised scale) between groups. Non-randomized comparative study provides low certainty evidence.

*Considered the number of participants relevant to the questions of this report, not all people included in the studies

Table 7: Characteristics of eligible non- comparative observational studies

Study ID	Sample size	Intervention	Outcomes measured	Likely to change conclusions	Reasons
Bordas, 2021	813	FtM bottom surgery	Surgical complications, satisfaction	No	Reports rate of complications (10.5%) and satisfaction (79% totally satisfied, 20% mainly satisfied) within range of effects reported by studies already included in systematic reviews. Unlikely to reduce imprecision and inconsistency within body of evidence (3177 and 1458 people, respectively) of non-comparative studies (42 and 27, respectively) to increase certainty of evidence
Elias, 2022	110	FtM top surgery	Complications	No	Reports rate of complications (16%) and revision surgery (5%), which is consistent with the rates reported in the studies included. Unlikely to increase the certainty of evidence

Effects of gender affirming therapies in people with gender dysphoria: evaluation of the best available evidence. Dr. Romina Brignardello-Petersen and Dr. Wojtek Wiercioch; Results; May 16, 2022

Figure 6: Risk of bias judgements for comparative studies

Study ID	Intervention	Confounding	Classification of the intervention	Deviations from intended interventions	Missing data	Measurement of outcome	Overall
Becker-Hebly, 2021	Puberty blockers, cross-sex hormones, or surgery						CRITICAL
Castelo-Branco, 2021	Cross-sex hormones						CRITICAL
Fontanari, 2020	Cross-sex hormones, cross-sex hormones or surgery						CRITICAL
Grannis, 2021	Cross-sex hormones						CRITICAL
Green, 2021	Cross-sex hormones						CRITICAL
Tordoff, 2022	Puberty blockers, cross-sex hormones						MODERATE
Turban, 2022	Cross-sex hormones						CRITICAL
Van Der Miesen, 2020	Puberty blockers						SERIOUS

Figure legend:

Low
Moderate
Serious
Critical
Unclear

App. 122

Effects of gender affirming therapies in people with gender dysphoria: evaluation of the best available evidence. Dr. Romina Brignardello-Petersen and Dr. Wojtek Wiercioch; Results; May 16, 2022

Figure 7: Risk of bias judgements for non-comparative studies

Study ID	Intervention	Representativeness of sample	Classification of intervention	Deviation from intended interventions	Missing data	Measurement of outcome	Overall
Bordas, 2021	FtM bottom surgery						LOW
Elias, 2022	FtM top surgery						MODERATE

Figure legend:

Low

Moderate

High

References

1. Ramos GGF, Mengai ACS, Daltro CAT, et al. Systematic Review: Puberty suppression with GnRH analogues in adolescents with gender incongruity. *Journal of endocrinological investigation* 2021;44(6):1151-58. doi: https://dx.doi.org/10.1007/s40618-020-01449-5
2. Rew L, Young CC, Monge M, et al. Review: Puberty blockers for transgender and gender diverse youth-a critical review of the literature. *Child and adolescent mental health* 2021;26(1):3-14. doi: https://dx.doi.org/10.1111/camh.12437
3. Excellence NIfHaC. Evidence review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria, 2020.
4. Quality AfHRa. Topic Brief: Treatments for Gender Dysphoria in Transgender Youth, 2021.
5. Baker KE, Wilson LM, Sharma R, et al. Hormone Therapy, Mental Health, and Quality of Life Among Transgender People: A Systematic Review. *Journal of the Endocrine Society* 2021;5(4):bvab011. doi: 10.1210/jendso/bvab011
6. Fledderus AC, Gout HA, Ogilvie AC, et al. Breast malignancy in female-to-male transsexuals: systematic review, case report, and recommendations for screening. *Breast (Edinburgh, Scotland)* 2020;53(9213011):92-100. doi: https://dx.doi.org/10.1016/j.breast.2020.06.008
7. Haupt C, Henke M, Kutschmar A, et al. Antiandrogen or estradiol treatment or both during hormone therapy in transitioning transgender women. *The Cochrane database of systematic reviews* 2020;11:CD013138. doi: 10.1002/14651858.CD013138.pub2
8. Karalexi MA, Georgakis MK, Dimitriou NG, et al. Gender-affirming hormone treatment and cognitive function in transgender young adults: a systematic review and meta-analysis. *Psychoneuroendocrinology* 2020;119:104721. doi: 10.1016/j.psyneuen.2020.104721
9. Kotamarti VS, Greige N, Heiman AJ, et al. Risk for Venous Thromboembolism in Transgender Patients Undergoing Cross-Sex Hormone Treatment: A Systematic Review. *The journal of sexual medicine* 2021 doi: 10.1016/j.jsxm.2021.04.006
10. Mattawanon N, Charoenkwan K, Tangpricha V. Sexual Dysfunction in Transgender People: A Systematic Review. *The Urologic clinics of North America* 2021;48(4):437-60. doi: 10.1016/j.ucl.2021.06.004
11. Totaro M, Palazzi S, Castellini C, et al. Risk of Venous Thromboembolism in Transgender People Undergoing Hormone Feminizing Therapy: A Prevalence Meta-Analysis and Meta-Regression Study. *Frontiers in endocrinology* 2021;12:741866. doi: 10.3389/fendo.2021.741866
12. Excellence NIfHaC. Evidence review: Gender-affirming hormones for children and adolescents with gender dysphoria., 2020.
13. Eftekhar Ardebili M, Janani L, Khazaei Z, et al. Quality of life in people with transsexuality after surgery: a systematic review and meta-analysis. *Health and quality of life outcomes* 2020;18(1):264. doi: 10.1186/s12955-020-01510-0
14. Bustos VP, Bustos SS, Mascaro A, et al. Regret after Gender-affirmation Surgery: A Systematic Review and Meta-analysis of Prevalence. *Plastic and reconstructive surgery Global open* 2021;9(3):e3477. doi: 10.1097/GOX.0000000000003477
15. Oles N, Darrach H, Landford W, et al. Gender Affirming Surgery: A Comprehensive, Systematic Review of All Peer-reviewed Literature and Methods of Assessing Patient-centered Outcomes (Part 2: Genital Reconstruction). *Annals of surgery* 2022;275(1):e67-e74. doi: 10.1097/SLA.0000000000004717
16. Oles N, Darrach H, Landford W, et al. Gender Affirming Surgery: A Comprehensive, Systematic Review of All Peer-Reviewed Literature and Methods of Assessing Patient-Centered Outcomes (Part 1: Breast/Chest, Face, and Voice). *Annals of surgery* 2022 doi: 10.1097/SLA.0000000000004728
17. Bustos VP, Bustos SS, Mascaro A, et al. Transgender and Gender-nonbinary Patient Satisfaction after Transmasculine Chest Surgery. *Plastic and reconstructive surgery Global open* 2021;9(3):e3479. doi: 10.1097/GOX.0000000000003479
18. Bustos SS, Bustos VP, Mascaro A, et al. Complications and Patient-reported Outcomes in Transfemale Vaginoplasty: An Updated Systematic Review and Meta-analysis. *Plastic and reconstructive surgery Global open* 2021;9(3):e3510. doi: 10.1097/GOX.0000000000003510
19. Tay YT, Lo CH. Use of peritoneum in neovagina construction in gender-affirming surgery: A systematic review. *ANZ journal of surgery* 2021 doi: 10.1111/ans.17147

20. 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. *European child & adolescent psychiatry* 2021;30(11):1755-67. doi: https://dx.doi.org/10.1007/s00787-020-01640-2
21. Castelo-Branco C, RiberaTorres L, Gomez-Gil E, et al. Psychopathological symptoms in Spanish subjects with gender dysphoria. A cross-sectional study. *Gynecological endocrinology : the official journal of the International Society of Gynecological Endocrinology* 2021;37(6):534-40. doi: https://dx.doi.org/10.1080/09513590.2021.1913113
22. Fontanari AMV, Vilanova F, Schneider MA, et al. Gender Affirmation Is Associated with Transgender and Gender Nonbinary Youth Mental Health Improvement. *LGBT health* 2020;7(5):237-47. doi: https://dx.doi.org/10.1089/lgbt.2019.0046
23. Grannis C, Leibowitz SF, Gahn S, et al. Testosterone treatment, internalizing symptoms, and body image dissatisfaction in transgender boys. *Psychoneuroendocrinology* 2021;132(7612148, qgc):105358. doi: https://dx.doi.org/10.1016/j.psyneuen.2021.105358
24. Green AE, DeChants JP, Price MN, et al. Association of Gender-Affirming Hormone Therapy With Depression, Thoughts of Suicide, and Attempted Suicide Among Transgender and Nonbinary Youth. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine* 2022;70(4):643-49. doi: https://dx.doi.org/10.1016/j.jadohealth.2021.10.036
25. Tordoff DM, Wanta JW, Collin A, et al. Mental Health Outcomes in Transgender and Nonbinary Youths Receiving Gender-Affirming Care. *JAMA network open* 2022;5(2):e220978. doi: https://dx.doi.org/10.1001/jamanetworkopen.2022.0978
26. Turban JL, King D, Kobe J, et al. Access to gender-affirming hormones during adolescence and mental health outcomes among transgender adults. *PloS one* 2022;17(1):e0261039. doi: https://dx.doi.org/10.1371/journal.pone.0261039
27. van der Miesen AIR, Steensma TD, de Vries ALC, et al. Psychological Functioning in Transgender Adolescents Before and After Gender-Affirmative Care Compared With Cisgender General Population Peers. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine* 2020;66(6):699-704. doi: https://dx.doi.org/10.1016/j.jadohealth.2019.12.018
28. Bordas N, Stojanovic B, Bizic M, et al. Metoidioplasty: Surgical Options and Outcomes in 813 Cases. *Frontiers in endocrinology* 2021;12(101555782):760284. doi: https://dx.doi.org/10.3389/fendo.2021.760284
29. Elias N, Rysin R, Kwartin S, et al. Breaking the Binary: The Approach to Chest Masculinizing Gender-Affirming Surgery in Trangender Men. *The Israel Medical Association journal : IMAJ* 2022;24(1):20-24.

ID	Study	Reason
#534	Abu-Ghname 2020	Wrong population: non transgender men
#434	Aires 2022	Wrong interventions: Other type of surgery: glottoplasty
#514	Angus 2021	Wrong outcomes: It does not include any outcome of interest. Includes: serum total testosterone concentration, body fat redistribution, breast development, and facial/body hair reduction
#318	Baddredine 2022	Wrong intervention. Continuing vs stopping estrogen during perioperative period of vaginoplasty
#40	Baram 2019	Wrong outcomes: only clinical outcomes are sperm count, testicular histology, hormone levels, etc.
#145	Barcelos 2022	Wrong outcomes: sexual satisfaction, desire, and function outcomes only
#60	Boczar 2021	No outcome data
#386	Bouman 2014	Wrong population: unclear that more than 80% are transgender
#208	Bustos 2021	Wrong intervention: niple areola reconstruction
#54	Connelly 2021	Wrong outcomes: Blood pressure
#43	Coon 2022	Wrong intervention: facial gender surgery
#34	D'Angelo 2018	Wrong design: narrative review
#165	Delgado-Ruiz 2019	Wrong outcomes: bone density
#355	Escandon 2022	Other type of surgery: facial surgery
#129	Fighera 2019	Wrong outcomes: bone mass
#597	Hembree 2017	Practice guideline, does not report the methods/ results of the systematic review in details
#120	Kakadekar 2021	Wrong outcomes: histological findings
#451	Kennedy 2021	Wrong intervention: self administered hormones
#375	Kloer 2021	Wrong outcomes: sexual health and satisfaction outcomes only
#439	Kovar 2019	More than 20% participants did not have gender dysphoria
#297	Kristensen 2021	Wrong outcomes: agression and hostility
#637	Leclere 2015	Wrong design: commentary of a systematic review
#293	Miranda 2021	Published in abstract format only
#624	Morrison 2016	Wrong intervention: facial feminization surgery
#270	Narayan 2021	Wrong design: narrative review
#119	Nolan 2019	Wrong intervention: phonosurgery
#167	Patel 2021	Wrong intervention: facial hair transplantation
#287	Ray 2020	Wrong population: cisgender is the population of interest, transgender included as indirect evidence and not in a systematic manner
#518	Rozga 2020	Published in abstract format only
#265	Sariyaka 2017	Wrong population: More than 20% participants did not have gender dysphoria
#35	Sayegh 2019	Wrong intervention: facial masculinization surgery
#124	Schwarz 2017	Wrong intervention: laryngeal surgery

#97	Siringo 2021	Wrong intervention: facial feminization surgery
#253	Song 2016	Wrong intervention: phonosurgery
#250	Song 2017	Wrong intervention: phonosurgery
#104	Spanos 2020	Wrong outcomes: lean mass, fat mass or insulin resistance
#257	Therattil 2017	Wrong intervention: thyroid chartilage reduction surgery
#328	Tirrell 2022	Wrong intervention: facial feminization surgery
#676	Traish 2010	Wrong design: narrative review
#279	VanDamme 2017	Wrong intervention: voice pitch raising surgery
#171	Vellho 2017	Wrong outcomes: BMI, blood pressure, hematocrit, hemoglobin, lipid profile, and liver enzymes
#112	Wilson 2020	Wrong outcomes: prolactine related outcomes (levels, hyperprolactinemia, prolactinoma)
#245	Worth 2018	Unable to access full text
#122	Ziegler 2018	Wrong outcomes: voice parameters and satisfaction with voice
#499	Zucker 2021	Unable to access full text

ID	Study	Reason
#1458	Al-Tamimi 2019	Wrong patient population
#287	Al-Tamimi 2020	Wrong study design: non comparative
#403	Alcon 2021	Wrong study design: non comparative
#214	Aldridge 2021	Wrong study design: non comparative
#54	Almazan 2021	Wrong patient population
#1387	Boas 2019	Wrong patient population
#1323	Branstrom 2020	Wrong patient population
#1447	Breidenstein 2019	Wrong study design: non comparative
#114	Briles 2022	Insufficient Sample Size <100
#1804	Butler 2019	Wrong patient population
#716	Carmichael 2021	Wrong study design: non comparative
#622	Cocchetti 2021	Wrong outcomes
#1067	Coon 2020	Wrong patient population
#1835	Cristofari 2019	Wrong patient population
#1486	Cuccolo 2019	Wrong patient population
#1276	deBlok 2020	Wrong patient population
#577	deRooij 2021	Wrong patient population
#1625	DeWolf 2019	Wrong patient population
#1759	Djordjevic 2019	Wrong patient population
#244	Falcone 2020	Insufficient Sample Size <100
#258	FosterSkewis 2021	Wrong comparator
#1583	Gallagher 2019	Wrong patient population
#139	Gumussoy 2022	Wrong study design: non comparative
#515	Hisle-Gorman 2021	Wrong study design: non comparative
#350	Hougen 2021	Insufficient Sample Size <100
#1007	Meyer 2020	Wrong study design: non comparative
#499	Miller 2021	Wrong patient population
#621	Mullins 2021	Wrong study design: non comparative
#1653	Naeimi 2019	Insufficient Sample Size <100
#1691	Namba 2019	Insufficient Sample Size <100
#1770	Neuville 2019	Insufficient Sample Size <100
#623	Neuville 2021	Insufficient Sample Size <100
#644	Nieder 2021	Insufficient Sample Size <100
#1624	Nikkels 2019	Wrong patient population
#353	Opsomer 2021	Wrong patient population
#1306	Papadopulos 2020	Wrong comparator
#640	Papadopulos 2021	Insufficient Sample Size <100
#1472	Pigot 2019	Wrong patient population
#899	Pigot 2020	Insufficient Sample Size <100
#1212	Segev-Becker 2020	Insufficient Sample Size <100
#1351	Staples 2020	Wrong outcomes
#645	Staud 2021	Insufficient Sample Size <100
#864	Terrier 2020	Insufficient Sample Size <100
#1083	vanderSluis 2020	Insufficient Sample Size <100

#1204	Veerman 2020	Insufficient Sample Size <100
#1409	Watanabe 2019	Wrong patient population
#512	Waterschoot 2021	Insufficient Sample Size <100