# Effects of puberty suppression in young people (under 18 years) experiencing gender incongruence or dysphoria: systematic review and meta-analysis

FOR PUBLIC CONSULTATION

17 October 2026

We acknowledge that the data in the studies included in this review represent the lived and living experience of trans, gender diverse and gender questioning young people and those who care for them, and that the content of this review addresses sensitive topics. As a review team, we recognise that there have been multiple evaluations of the forms of health care provided to individuals seeking to affirm their gender. While replication of research is an essential part of science (1), and decision-makers seek the best and most relevant evidence for their context, we acknowledge the impact that this scale of inquiry may have.

We are especially grateful to the people with lived experience, members of the Gender Guidelines Development Committee and members of the Gender Guidelines Reference Panel who have generously shared their expertise and experience to ensure that this review is both relevant and respectful to those affected by the decisions that it will inform. A statement of our positionality is provided at the end of this report.

Authors: Sue E Brennan[1]*, Steve McDonald[1], Simon Turner[2], Matthew J Page[2], David Clark[2] and Joanne E McKenzie[2]

[1] Cochrane Australia, School of Public Health and Preventive Medicine, Monash University

[2] Methods in Evidence Synthesis Unit, School of Public Health and Preventive Medicine, Monash University

* Contact author (sue.brennan@monash.edu)

# Contents

Abbreviations used in this report ..... 2
1 Background ..... 4
2 Objectives ..... 6
3 Methods ..... 7
Involvement of interest holders and people with lived experience ..... 7
Terminology ..... 8
3.1 Criteria for considering studies for this review ..... 8
3.1.1 Types of studies ..... 8
3.1.2 Types of participants ..... 11
3.1.3 Types of interventions ..... 11
3.1.4 Types of outcomes ..... 11
3.2 Search methods for identification of studies ..... 14
3.2.1 Published systematic reviews and evidence reports ..... 14
3.2.2 Database searches ..... 16
3.2.3 Register searches ..... 16
3.2.4 Searching other sources ..... 16
3.3 Data collection and analysis ..... 17
3.3.1 Selection of studies ..... 17
3.3.2 Data extraction and management ..... 17
3.3.3 Assessment of risk of bias ..... 20
3.3.4 Measures of treatment effect ..... 22
3.3.5 Unit of analysis issues ..... 23
3.3.6 Dealing with missing data ..... 23
3.3.7 Assessment of heterogeneity ..... 23
3.3.8 Assessment of biases due to missing results and integrity ..... 23
3.3.9 Data synthesis ..... 24
3.3.10 Summary of findings tables and assessment of certainty of the evidence ..... 25
3.3.11 Interpretation of findings (evidence statements) ..... 27
4 Results ..... 27
4.1 Search results ..... 27
4.2 Description of studies ..... 29
4.2.1 Study setting and design ..... 29
4.2.2 Characteristics of those receiving puberty suppressants ..... 30
4.2.3 Puberty suppression interventions ..... 30
4.2.4 Comparators ..... 31

FOR PUBLIC CONSULTATION

4.2.5 Outcomes ..... 32
4.3 Effects of puberty suppression ..... 32
4.3.1 Summary of findings ..... 32
4.3.2 Critical outcomes ..... 37
4.3.4 Important outcomes ..... 44
5 Discussion ..... 49
5.1 Summary of main results ..... 49
5.2 Limitations of the evidence included in the review ..... 50
5.3 Limitations of the review processes ..... 51
5.4 Agreements and disagreements with other studies or reviews ..... 52
6 Authors' conclusions ..... 55
6.1 Implications for practice ..... 55
6.2 Implications for research ..... 55
6.2.3 Equity related implications for research ..... 56
7 Additional information ..... 57
7.1 Acknowledgements ..... 57
7.2 Contributions of authors ..... 57
7.3 Declarations of interest ..... 58
7.4 Sources of support ..... 58
Internal sources ..... 58
External sources ..... 58
7.5 Registration and protocol ..... 58
Data, code and other materials [availability of] ..... 59
Disclosure of artificial intelligence use ..... 59
Positionality statement ..... 59
8 References ..... 60
Appendices ..... 69
Appendix 1a. Outcome domains prioritised for this review ..... 69
Appendix 1b. Outcomes that were considered but not prioritised for the review ..... 77

FOR PUBLIC CONSULTATION

In October 2025 Monash University was contracted by the National Health and Medical Research Council (NHMRC) to design and undertake the systematic review described in this report. This systematic review is one of several independent contracted evidence evaluations being undertaken for the *National clinical practice guidelines for the care of trans and gender diverse people under 18* being developed by NHMRC. The review was designed in collaboration with the Office of NHMRC (ONHMRC), and in consultation with members of the Gender Guidelines Development Committee (the Committee) and the Gender Guidelines Reference Panel (the Reference Panel). The protocol for this review was approved by the NHMRC in March 2026.

## Copyright

All materials presented in this report are licensed under a Creative Commons Attribution 4.0 International Licence (CC BY 4.0) with the exception of content identified as being owned by third parties. You are free to share and adapt the material, provided you attribute the work (including any material sourced from it) to the authors and the National Health and Medical Research Council (NHMRC) using the suggested citation below. To view a copy of this licence, visit Deed - Attribution 4.0 International Creative Commons.

## Suggested citation (attribution):

Brennan SE, McDonald S, Turner S, Page MJ, Clark D, McKenzie JE. Effects of puberty suppression in young people (under 18) experiencing gender incongruence or dysphoria: systematic review and meta-analysis. Report prepared for the National Health and Medical Research Council, Canberra, Australia. October 2026.

## Use of images

Unless otherwise stated, all images (including background images, icons and illustrations) are copyrighted by their original owners.

## Contact

To obtain information regarding the Gender Guidelines, contact genderguidelines@nhmrc.gov.au.

FOR PUBLIC CONSULTATION

# Abbreviations used in this report

| Abbreviation | Full term / explanation |
|---|---|
| **aBMAD** | Areal bone mineral apparent density |
| **AFAB** | Assigned female at birth |
| **AMAB** | Assigned male at birth |
| **BMD** | Bone mineral density |
| **CI** | Confidence interval |
| **$CI_{upper}$** | Confidence interval upper limit |
| **$CI_{lower}$** | Confidence interval lower limit |
| **the Committee** | the Gender Guidelines Development Committee |
| **Correlation** | Correlation between the baseline and follow-up measure |
| **DSM** | Diagnostic and Statistical Manual of Mental Disorders |
| **GAC** | Gender affirming care |
| **GAH(T)** | Gender affirming hormone (therapy) |
| **GAS** | Gender affirming surgery |
| **GD** | Gender dysphoria |
| **GGDC** | Gender Guidelines Development Committee |
| **GGRP** | Gender Guidelines Reference Panel |
| **GnRHa** | Gonadotropin-releasing hormone (GnRH) agonist |
| **GRADE** | Grading of Recommendations Assessment, Development and Evaluation |
| **ICD** | International classification of diseases |
| **J** | Small-sample correction |
| **PS** | Puberty suppressants |
| **q1** | $25^{th}$ percentile |
| **q3** | $75^{th}$ percentile |
| **N/A** | Not applicable |
| **NHMRC / ONHMRC** | National Health and Medical Research Council / Office of the NHMRC |
| **N/R** | Not reported (used extensively in the study characteristics supplement) |
| **NRS / NRSI** | Non-randomised studies of interventions |
| **NSW** | New South Wales |
| **NT** | Northern Territory |
| **PBS** | Pharmaceutical Benefits Scheme |
| **PICO** | Population(s), Intervention(s), Comparator(s), Outcome(s) |
| **QLD** | Queensland |
| **QoL** | Quality of life |
| **RoB** | Risk of bias |

| Abbreviation | Full term / explanation |
|---|---|
| **RCT / RT** | Randomised controlled trials / randomised trials |
| **Reference Panel** | the Gender Guidelines Reference Panel |
| **ROBINS-I** | Risk of bias in non-randomised studies of interventions |
| **SAB** | Sex assigned at birth |
| **SD** | Standard deviation |
| **SMD** | Standardised mean difference |
| **SR** | Systematic review |
| **TGA** | Therapeutic Goods Administration |
| **UK** | United Kingdom |
| **USA / US** | United States of America |
| **WHO** | World Health Organization |
| **WPATH** | World Professional Association for Transgender Health |

FOR PUBLIC CONSULTATION

# 1 Background

As the nation's leading expert body in health and medical research, the National Health and Medical Research Council (NHMRC) has a statutory responsibility for developing and supporting high quality guidelines for clinical practice under the National Health and Medical Research Council Act (1992). On 31 January 2025 the Minister for Health and Ageing, the Hon Mark Butler, asked NHMRC to develop new national clinical practice guidelines for the care of trans and gender diverse people under 18 (Gender Guidelines)[1]. The Gender Guidelines will support clinicians and trans and gender diverse people under 18 across Australia by providing consistent, evidence-based recommendations that can be adapted to diverse service models and clinical contexts. This work reflects growing recognition of the need for consistent and evidence-based clinical guidance in this area.

While national prevalence data for trans and gender diverse young people in Australia are limited, survey findings provide some insight. A 2021 study reported that around 7% of adolescents aged 14–18 identified as trans or gender diverse (2), an increase from around 2% in a 2018 survey of the same age group (3). Trans and gender diverse young people experience ongoing health challenges and inconsistent access to care (4-7).

The Australian Standards of Care and Treatment Guidelines for Trans and Gender Diverse Children and Adolescents are the guidelines that are commonly used in clinical practice in Australia (8); however, these guidelines have not been approved by NHMRC.

The development of the Gender Guidelines will align with NHMRC Standards for Guidelines and international best practice in guideline development, and the use of GRADE methodology to assess certainty of evidence and balance benefits and harms. Recommendations will be informed by three targeted systematic evidence reviews. The first of these reviews examines the effects of puberty suppression in young people under 18 years with gender dysphoria or incongruence.

## Description of body image distress and gender dysphoria

Trans and gender diverse young people experience unjust and avoidable health disparities arising from stigma, discrimination, and barriers to accessing health care (4, 9, 10). These factors are associated with body image distress and gender dysphoria, which can have downstream consequences for wellbeing, mental health (including depression, anxiety, and suicidal thoughts and behaviours), and functioning at school, socially, and within families (10-13).

The term 'gender dysphoria' is generally used clinically to describe affective distress that "may be experienced because a person's gender identity differs from that which is physically and/or socially attributed to their sex assigned at birth" (WPATH, v8, S252[2]). In the Diagnostic and Statistical Manual of Mental Disorders, 5th edition, a diagnosis of gender dysphoria is based on "a marked incongruence between an individual's experienced/expressed gender and their assigned gender that has been present for at least 6 months and is associated with clinically significant distress or impairment in social, occupational, or other important areas of

[1] https://www.nhmrc.gov.au/health-advice/guidelines-care-trans-and-gender-diverse-people

[2] World Professional Association for Transgender Health. Standards of Care Version 8 wpath.org/publications/soc8/.

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License





FOR PUBLIC CONSULTATION

functioning" (14). Concepts grounded in the lived experience of trans and gender diverse people (15) consider the influence of exposure to minority stress (stemming from stigma, discrimination, non-acceptance, etc.) on body dysphoria (arising from physical characteristics) (11, 12, 15, 16). Not all trans and gender diverse people experience gender dysphoria, and not all trans and gender diverse young people who access care for gender dysphoria seek medical intervention (including puberty suppression or gender affirming hormones) (12).

### Description of puberty suppression and how it might work

Puberty suppressants are medications that are used to pause endogenous sex hormone production in puberty and halt the physical changes that these hormones induce (e.g. voice deepening, hair growth, breast development, final height, body habitus, and menstruation) (8, 17). For trans and gender diverse young people, body image distress and gender dysphoria may intensify with the onset of puberty, as unwanted physical changes (secondary sex characteristics) emerge in response to endogenous sex hormones (18, 19). By pausing pubertal development, puberty suppressants may help stabilise body image distress and gender dysphoria in the short term, while allowing time to clarify embodiment goals and to consider whether to proceed with gender affirming medical treatment (18, 19).

### Initiation of puberty suppression

In Australia, puberty suppressants are typically prescribed by specialists such as paediatricians or paediatric endocrinologists following comprehensive multidisciplinary assessment in tertiary healthcare settings. Initiation of puberty suppression usually occurs in early to mid-puberty (Tanner stages 2–3), although it may be used in later puberty for specific indications such as menstrual suppression. Treatment duration with puberty suppressants alone is typically 2–3 years (if not discontinued earlier). Individuals may or may not transition to gender affirming hormones (e.g. testosterone, oestrogen).

Provision of gender services for young people in Australia is led by states and territories, who are responsible for relevant services. Young people prescribed puberty suppression as part of gender affirming care in Australia are generally expected to meet specific clinical criteria, including a diagnosis of persistent gender dysphoria and evidence that gender incongruence is experienced and sustained over time, particularly where distress is exacerbated by the onset or progression of puberty. Legal requirements must also be met, including agreement from the young person and informed consent from relevant legal guardians, with family court orders required if guardian consent is not obtained.[3]

### Puberty suppressing medications available in Australia

Three gonadotropin-releasing hormone (GnRH) agonist medications are currently used for puberty suppression in Australia: leuprorelin (Lucrin Depot, Eligard), goserelin (Zoladex) (implant formulation), and triptorelin (Diphereline, Decapeptyl SR). All prescribing of GnRH agonists for gender dysphoria in Australia is off-label, which means that a medicine is used for an indication, at a dose, via a route of administration or for a patient group that

[3] Legal requirements as to informed consent vary across Australian jurisdictions. In Queensland, a Health Service Directive (HSD) was issued on 28 January 2025, temporarily pausing puberty suppression and gender affirming hormones for new patients under 18 with gender dysphoria.

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License


FOR PUBLIC CONSULTATION

is not included in the approved Therapeutic Goods Administration (TGA) product information (20). Off-label prescribing is common, with rates estimated to be up to 90% in paediatric patients (21, 22). As a result of off-label use, GnRHa medications are not subsidised under the Pharmaceutical Benefits Scheme (PBS). However, if prescribed in a public hospital setting (or under the supervision of a clinician in this setting), the public hospital system of each state and territory will cover the cost (apart from QLD and NT where care is no longer provided for new patients in the public hospital system).

### Why it is important to do this review

Multiple systematic reviews have examined the effects of puberty suppression for young people experiencing gender incongruence or dysphoria (23-34). These evidence reviews are published in the peer-reviewed research literature and reports commissioned by various government agencies around the world. Existing evidence reviews vary widely in terms of the questions addressed and review methodology. The evidence reviews conducted for jurisdictions in Australia (specifically New South Wales Health and Queensland Health) included data from existing systematic reviews, supplemented where necessary with primary studies (i.e. they used 'overviews of reviews' methods) (25, 30). Although this enables rapid review of evidence by capitalising on existing synthesis, there are trade-offs in terms of the match between the questions of interest and those addressed in each review, reliance on the accuracy of data available in reviews, and limitations in review methodology and reporting (35). These limitations can arise whenever an existing evidence review is used. Very few systematic reviews of primary studies on puberty suppression adhere to Cochrane or GRADE (Grading of Recommendations Assessment, Development and Evaluation) methods (Miroshnychenko is an exception (29)[4]), and none of the reviews reported involvement of people with lived experience and other interest holders in the prioritisation of outcomes. These are important limitations for guideline development that are addressed in this new review.

# 2 Objectives

The objective of this evidence review is to examine the effects of puberty suppressing medications in young people experiencing gender incongruence or dysphoria on

- critical outcomes for decision-making[5]: pubertal development, gender dysphoria, body image and satisfaction, serious adverse effects, and
- important outcomes for decision-making: quality of life and wellbeing, functioning, mental health, satisfaction with gender affirming care and adverse effects (bone health and fertility).

FOR PUBLIC CONSULTATION

[4] Concerns have been raised about the sponsors of this review and the implications of their involvement in developing the review question and as authors of the protocol (PROSPERO CRD42023452171). In response, the review authors issued a statement (https://hei.healthsci.mcmaster.ca/systematic-reviews-related-to-gender-affirming-care/).

[5] Outcomes to be considered in the evidence review were those prioritised as critical or important by the Committee (see section 3.1.4 Outcome prioritisation process).

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

# 3 Methods

Methods reported in this review are based on the *Cochrane Handbook for Systematic Reviews of Interventions* (36). The GRADE approach was used to prioritise outcomes, interpret findings and assess the certainty of evidence arising from this review (see Sections 3.1.4, 3.3.4 and 3.3.10 for details). The review is reported in accordance with the PRISMA 2020 statement (37, 38). The methods were prespecified in the protocol for the review (not registered on PROSPERO), which underwent independent methodological review (contracted by the NHMRC) and peer review by the Committee and the Reference Panel (see Supplement 1, Appendix S1.E for a summary of comments and responses). Any changes to the methodology (e.g. to eligibility criteria) were finalised before screening of studies commenced. The protocol was approved by the NHMRC, and the final version posted to the Bridges repository (see protocol here). Any difference between the protocol and the review (including methods not used) are documented in Supplement 1, Appendix S1.D.

## Involvement of interest holders and people with lived experience

The protocol for this review was developed in consultation with the NHMRC, the Gender Guidelines Development Committee ('the Committee') and members of the Gender Guidelines Reference Panel ('the Reference Panel'). The Committee comprises 18 members, including clinicians, researchers, and four individuals with lived experience of accessing gender services in Australia, either personally as young people (including those who are trans and gender diverse) or as someone caring for a person who has. The Reference Panel includes young people who have accessed gender services in Australia, parents, caregivers, and clinicians who provide care to trans and gender diverse young people (see governance structure at https://www.nhmrc.gov.au/health-advice/guidelines-care-trans-and-gender-diverse-people/governance).

Interest holder contributions to the protocol included identification and prioritisation of outcomes (see 'Types of outcomes'), advice that led to the underpinning logic model for describing the anticipated effects from puberty suppression, advice on review eligibility criteria, and advice on background literature and terminology. The resulting draft protocol was circulated by the NHMRC to the Committee and the Reference Panel for critical review (24 people provided comments; 9 from the Committee and 14 from the Reference Panel).

Interest holder contributions to the evidence review included confirming thresholds for interpreting findings and advice on optimising presentation and interpretation of results based on lived and clinical experience. The results of the review were presented twice to the Committee prior to the meetings at which the GRADE evidence to decision process was conducted and recommendations were made. The proposed thresholds for interpreting findings were presented and agreed at the first Committee meeting (March 2026) and then results from the preliminary synthesis were shared for the first time. The Committee's advice on presentation and interpretation of findings were integrated into the review prior to presenting the findings at the second Committee meeting (May 2026) at which results were presented in a simulated GRADE Evidence to Decision process. Prior to each meeting, drop-in sessions were held for lived experience members from the Committee to discuss the GRADE approach, and the presentation and interpretation of findings. Prior to the recommendations meeting, drop-in sessions were held with members of the Committee that had been unable to attend the second presentation of results. Through these processes, we aimed to ensure that all members of the Committee had maximum opportunity to consider

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

data from the evidence review, and for the evidence review team to receive advice on optimising the presentation and interpretation of findings.

### Terminology

In this report, we use the terminology adopted for the Gender Guidelines and aim to standardise language to facilitate comparison across studies. Throughout the review we use the term 'participants', which is a common convention in systematic reviews, however, we acknowledge that much of the data comes from routine care (i.e. is observational), rather than from a context in which there is formal recruitment of participants to a study.

We are aware that using the terminology of included studies brings the risk of perpetuating language that is outdated and pathologising. This concern has been raised in relation to some commonly used scales for measuring outcomes (16, 39). When reporting characteristics of included studies (such as in the characteristics of included studies table and risk of bias assessments), we use terms reported in the included study when doing otherwise would require assumptions. Any exceptions to this approach are indicated. Of the studies that reported gender, most did so as a binary (e.g. trans girl/woman, trans boy/man). In these instances, we could not be sure whether gender was self-reported or assumed by the study authors. Non-binary gender identities and gender fluidity were rarely reported in the studies.

## 3.1 Criteria for considering studies for this review

### 3.1.1 Types of studies

Randomised controlled trials (RCTs) and non-randomised studies of interventions (NRSIs) were eligible for inclusion in the review.

*Randomised controlled trials (RCTs)* including individually and cluster randomised were eligible for inclusion in the review. However, we did not identify any randomised trials and recognise that ethical and feasibility considerations have been identified as precluding trials of puberty suppression because such trials involve withholding treatment thus leading to permanent physical changes.

*Non-randomised studies of interventions (NRSIs)* with specific design features were eligible for inclusion in the review. We were deliberately inclusive of a range of study designs to provide a comprehensive inventory of the available studies (and their design features). However, the strength of the study designs for inferring causality differs, which is examined through our synthesis (see 'Data synthesis'). Study design labels are used as an aid for communicating about eligible designs (Table 1); however, eligibility decisions were based on assessment of the specific design features of each study rather than the label used by the study authors (Table 1) (40, 41). The eligibility of the design was assessed on an outcome basis (see exclusion criteria below).

Eligible designs are those in which the following features are present.

- Intervention may be allocated to individuals or clusters. In principle, this could include allocation by cluster (e.g. implementation of a clinic-wide policy to offer puberty suppressants to all patients). However,

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

such allocation is unlikely in practice, given the philosophy of individualised gender affirming care. For this reason, we do not describe cluster designs in **Table 1**.

- Intervention groups may be formed by some action of the researchers, or in the course of usual treatment decisions (including healthcare decision-makers, practitioners or participants/patients/peoples' choices) (42).
- The population of the comparator group may be:
    - a group of individuals who do not receive puberty suppressants and are the same population (trans and gender diverse young people)
    - the baseline measure(s) of the outcome for young people who do receive puberty suppressants (i.e. an uncontrolled before-after study) for critical outcomes.
- There must be an attempt to control for confounding either by design (e.g. matching individuals who receive and do not receive puberty suppressants on particular characteristics (e.g. age, sex assigned at birth), collecting outcome measure(s) prior to receipt of puberty suppressants) or through statistical analyses that adjust for confounders.

**Table 1**. Study design labels used in the review with indicative features

| Study design label | Thumbnail sketch of indicative features (adapted from (33, 34)) |
|---|---|
| Non-randomised controlled trial (NRCT) | Study in which allocation to intervention and comparator is not random or quasi-random (including where based on the practitioners' or participants' preferences) and is applied in the context of a research study. Participants are followed up from the start of intervention up to a later time for ascertainment of outcomes of interest (41, 43). In practice this design is unlikely to be materially different from a prospective cohort study. |
| Prospective cohort study (PCS) | Study in which participants are identified prospectively and classified as having received the intervention or comparator of interest on the basis of the prospectively collected information. Participants are followed up from the start of intervention up to a later time for ascertainment of outcomes of interest (41, 43). |
| Retrospective cohort study (RCS) | Study in which participants are identified from historic records and classified as having received the intervention or comparator of interest on the basis of the historic information. Participants are followed up from the start of intervention up to a later time for ascertainment of outcomes of interest (41, 43). |
| Target trial emulation analysis | Analysis of a cohort using the 'target trial' framework. This involves two steps: first, defining the causal question of interest in terms of a protocol for a hypothetical randomised trial (the 'target trial'); and, second, emulating the components of the protocol using the observational data (44). |

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

| Study design label | Thumbnail sketch of indicative features (adapted from (33, 34)) |
|---|---|
| Case-control study | Study in which individuals experiencing an outcome of interest are identified from within a defined population and form a group of 'cases'. Individuals, who may be matched to the cases, who did not experience the outcome of interest are also identified from within the defined population and form the group of 'controls' (41). |
| Uncontrolled before-after studies (including interrupted time series (ITS)) | Study that collects longitudinal data on individuals both before and after implementation of the intervention. Studies in which there are several measurement times before and after are ITS, and are also referred to as repeated measures studies (43). |
| Controlled before-after studies (including controlled interrupted time series (CITS)) | As for uncontrolled before-after studies and ITS but with a control group of participants who do not receive the intervention (43). |
| Cross-sectional study | Study in which information required to classify individuals according to receipt of the intervention or not and according to outcome are collected at the same time, but where it can be established that the intervention preceded the outcome (41). |

**Exclusions**. Cross-sectional studies in which it cannot be established if puberty suppressants preceded the outcome because a causal effect cannot be established. Case series (i.e. a single cohort with no outcome measure before implementation of the intervention). This includes studies labelled by the primary investigators as before-after but for which an outcome can only be measured after receipt of the intervention (e.g. death by suicide). In practice these are case series because there is no comparator; there is only post intervention data for the group receiving the intervention. Before-after studies that do not report a critical outcome.

Studies must be peer reviewed (except for preprints) and published in full (in any language). Studies published only as conference abstracts or theses were not eligible for the review (except where a full-text report was available) but are reported in the list of studies excluded, ongoing or awaiting assessment (Supplement 2) The decision to exclude conference abstracts from data extraction reflects the complexity of evaluating study design and interpreting results with such limited information. In general, it is not possible to confirm whether the study uses an eligible design from an abstract alone because we use study design features rather than labels to determine eligibility. In addition, essential information about the analysis and interpretation of results is rarely provided and there is heightened potential for authors to have selectively reported the most favourable results in an abstract.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

### 3.1.2 Types of participants

We included studies involving people who initiated puberty suppression before 18 years of age as part of medical care for gender incongruence (historically referred to in the ICD-10 as 'gender identity disorder', a classification now deprecated (45)) with or without a diagnosis of gender dysphoria. Where a large majority were likely to have received puberty suppression before age 18, we considered the study eligible and assessed the implications when rating the certainty of evidence (indirectness), since this approach maximises available data while taking account of the applicability of findings (46). Participants could be of any age at follow-up (to capture longer-term and adverse outcomes).

### 3.1.3 Types of interventions

We included studies of gonadotropin-releasing hormone (GnRH) agonist medications for suppressing puberty (sometimes referred to as 'puberty suppressants' or 'puberty blockers'). Any GnRH agonist prescribed in Australia was eligible, including leuprorelin (Lucrin Depot, Eligard), goserelin (Zoladex) and triptorelin (Diphereline, Decapeptyl SR). Other GnRH agonists, such as histrelin implants were also eligible. In the draft protocol, we indicated that other anti-androgens or progestins (e.g. to suppress menstrual cycle) would be eligible, however we excluded these medications following feedback from the Reference Panel ahead of commencing formal screening of studies.

Any dosage or route of administration of puberty suppressants was eligible, noting that administration via either intramuscular injection or implant is usual (depending on the GnRHa). While there were no explicit limits on the duration of treatment, puberty suppressants are intended for temporary use.

#### Comparisons

1. Puberty suppressants versus no puberty suppressants (no intervention, a co-intervention offered to both groups, or continuation of usual care).

Any co-intervention was eligible (i.e. pharmacological or non-pharmacological) providing that the effects of puberty suppression could be isolated. This could include social and psychological care intended to support a person in their gender identity. Since gender affirming hormone (GAH) therapy brings about physical changes intended to affirm gender, we only considered inclusion of studies in which participants had received GAH in exceptional circumstances (as reported in results). Usual care comparators were eligible, meaning that participants could continue to access their routine care or therapy (i.e. any other healthcare, gender affirming care, or both). In practice, it was generally not be possible to identify the exact nature of the comparator.

As described under 'Types of studies', the comparator group could be either a different group of individuals or the same individuals with results compared over time (as in uncontrolled before-after studies).

### 3.1.4 Types of outcomes

Studies were only eligible for the review if they reported at least one outcome prioritised as critical or important for making recommendations in the guideline.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

### Outcome prioritisation methods

Outcomes for this review were prioritised through an iterative process, that was led by the NHMRC and adhered to methods recommended by GRADE and Cochrane (46, 47).

In brief, the process involved the following steps.

- Identification of a comprehensive list of potentially relevant outcomes from existing guidelines, systematic reviews and other sources.
- Grouping of related outcomes by domain and description of each domain.
- Iterative rounds of discussion, survey and feedback from the Committee and the Reference Panel (as above, comprised of people with lived experience and experts in the clinical and/or research field) to refine and prioritise the outcomes.
- Targeted consultation to elicit further expert advice from researchers, clinicians and people with lived experience (primarily, members of the Committee and the Reference Panel) and thus ensure judgements about the importance and relevance of outcomes were well-informed.
- Approval of the prioritised outcomes by the Committee prior to study selection or data extraction.

In prioritising outcomes, the focus was on outcomes of most importance to young people seeking care for gender incongruence and their families or caregivers. Outcomes categorised as critical for decision-making had to be directly relevant to the main treatment goals of puberty suppression (i.e. direct effects on the outcomes were anticipated). Other highly valued outcomes were included as important but not critical if effects were likely to be indirect (e.g. treating co-occurring depression, anxiety and suicidality is paramount, but puberty suppressants are not mental health treatments, and any effects are expected to be dependent on the outcomes puberty suppression targets directly in combination with mental health care).

### Outcomes prioritised for the review

Prioritised outcomes are listed below. **Figure 1** shows the underpinning logic for how these outcomes relate to the primary and secondary goals of treatment. Each outcome domain is described in **Appendix 1a** in a table reporting (1) the domain name, (2) the domain specification (definitions; inclusions, exclusions), (3) examples of eligible measures, (4) the rationale for inclusion of each domain (and any exclusions), and (5) the timeframe(s) for follow-up.

Critical outcomes for decision-making

1. Pubertal development (during treatment; after discontinuation of puberty suppressants)
2. Gender dysphoria (during treatment; end of treatment period)
3. Body image (during treatment; end of treatment period)
4. Serious adverse events: any serious untoward outcome that occurs during, or after, the use of puberty suppressants, but not necessarily caused by the puberty suppressant (short-term follow-up (during or end of treatment); longer-term if effects can be isolated from subsequent medical care).

Important outcomes for decision-making but not critical

5. Quality of life and wellbeing (end of treatment period)

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

6. Functioning (social, family; any context) (end of treatment period)
7. Satisfaction with care (end of treatment period)
8. Mental health: depression symptoms, anxiety symptoms, suicidal thoughts and behaviours (end of treatment period)
9. Mental health: death by suicide (end of treatment period)
10. Adverse events of special interest: bone health – fractures (during treatment; after cessation of treatment if effects can be isolated from other exposures, including subsequent treatment)
11. Adverse events of special interest: fertility – reproductive potential (end of treatment period).

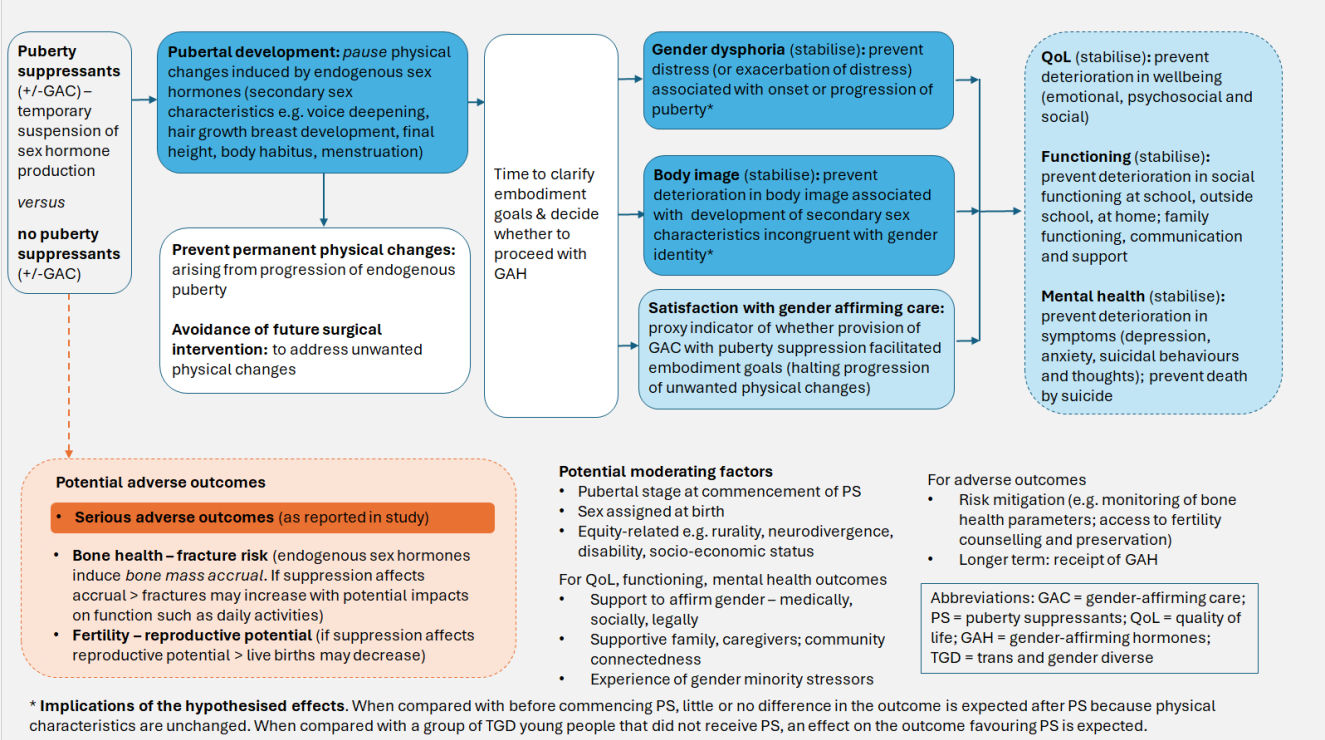



**Fig 1** | Logic model showing primary and secondary goals of treatment (dark and light blue cells respectively) and potential adverse outcomes. Darker shading (blue or orange) = critical outcomes for decision-making. Lighter shading = important but not critical outcomes for decision-making. Cells shaded in white are mediators (time) or a longer-term outcome not addressed in the review except where eligible for prioritised outcome domains. Sources informing the model include (12, 17, 19, 48)

We prespecified the anticipated effects (**Figure 1**), based on discussion at Committee meetings and targeted advice from those with lived or clinical experience from the Committee or Reference Panel. The resulting logic model shows the primary and secondary goals of treatment, potential adverse outcomes and effect modifiers (the latter derived from models of gender minority stress e.g. (12)) (49). Puberty suppression is used to pause progression of physical changes induced by endogenous puberty, but it does not change primary sex characteristics, nor does it bring about physical changes aligned with gender. Temporarily halting puberty may prevent body image distress (or worsening of distress) arising from decreasing gender congruence if puberty were to progress, but it is not expected to ameliorate gender dysphoria or downstream outcomes related to

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

psychosocial function and quality of life. This has implications for the interpretation of results; as noted in **Figure 1** little or no difference is expected if the outcome is compared before and after treatment within a group.

Outcomes that were considered but not prioritised for the puberty suppression review are tabulated in **Appendix 1b** with the rationale for why the outcomes were not prioritised. Other outcomes/concepts were considered as part of the contextual information used to develop the interim advice (reversibility, discontinuation, reidentification and regret).

**Selection of results**

From each study, we selected one result per outcome domain for data extraction, risk of bias assessment and reporting of results in the summary and synthesis. Selecting one result per study for inclusion in each analysis ensures that individual studies do not receive too much weight in a synthesis (meta-analysis or other synthesis). Where studies report multiple results, the result for inclusion in the synthesis was selected based on relevance and availability of data. In practice, although some studies reported multiple results that were eligible for a domain, the data were sometimes presented in a format that could not be used for meta-analysis leaving only one result for inclusion.

## 3.2 Search methods for identification of studies

Eligible studies were identified from two main sources:

1. Studies that were included in recent systematic reviews and evidence reports of adolescents with gender dysphoria/incongruence (section 3.2.1)
2. Search of several bibliographic databases (MEDLINE, Embase, PsycInfo, CINAHL and Europe PMC) covering the period January 2024 to 6 March 2026 (section 3.2.2)

### 3.2.1 Published systematic reviews and evidence reports

We created a repository of the studies included in recent systematic reviews and other evidence reports of puberty suppression for gender incongruence or dysphoria. The six systematic reviews and three evidence reports used to identify relevant studies (published between 2021 and 2025) are listed in **Table 2**. They included the systematic review by Taylor et al. that underpinned the 2024 Cass Review and three recent independent reviews commissioned by jurisdictions in Australia and New Zealand. Several of these reviews covered both puberty suppression and gender affirming hormone therapy, but only studies of puberty suppression (or puberty suppression plus gender affirming hormone therapy) were included for screening.

The most recent search was conducted in June 2025 (Queensland Health). Nearly all reviews included searches of MEDLINE, Embase and PsycInfo, but across the reviews, 10 other sources of evidence were also searched. The number of studies included per review ranged from 9 to 51, reflecting different inclusion criteria, particularly in relation to eligible study designs.

Collectively, these reviews included 100 unique primary studies (see Studies assessed tab of Supplement 2). Despite the wide range of databases and other sources searched, all but three of the studies were in MEDLINE, with the remaining in Embase or PsycInfo. In addition to screening these 100 studies, we also screened studies

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

excluded at full-text review from the systematic reviews by Miroshnychenko and Tornese (since these reviews provided comprehensive lists of exclusions).

**Table 2.** Systematic and other evidence reports of puberty suppression for gender incongruence or dysphoria used to generate the pool of potentially eligible studies as well as the number of studies included in each review

| Systematic reviews | Evidence sources | Search date | Studies included in each review |
|---|---|---|---|
| Rew 2021 (31) | PubMed, PsycInfo, LGBT Life and Web of Science | Not reported | 9 |
| Thompson 2022 (33) | MEDLINE, Embase, PsycInfo and CINAHL. PROSPERO and Cochrane for SRs. | 2020 Nov | 19 |
| Ludvigsson 2023 (28) | PubMed, Embase, PsycInfo, CINAHL, Cochrane Library, Scopus and SocINDEX | 2021 Nov | 24 |
| Taylor 2024 (32) | MEDLINE, Embase, PsycInfo, CINAHL and Web of Science (Social Science Citation Index) | 2022 Apr | 50 |
| Miroshnychenko 2025 (29) | MEDLINE, Embase, PsycInfo, Social Sciences Abstracts, Contemporary Women's Issues, LGBTQ+Source, Sociological Abstracts, Studies on Women, Gender Abstracts and Google Scholar | 2023 Sep | 10 |
| Tornese 2025 (34) | PubMed, Embase, Cochrane Library, Web of Science, ClinicalTrials.gov and WHO International Clinical Trials Registry Platform | 2024 Feb | 51 |
| Australian / NZ evidence reviews | | | |
| NSW Ministry of Health (Sax) 2024 (25) | MEDLINE, Embase, PsycInfo, CINAHL, Scopus, Cochrane Library and JBI | 2023 Sep | 31 |
| NZ Ministry of Health 2024 (23, 24) | MEDLINE, Embase, PsycInfo and Scopus | 2024 May | 41 |
| Queensland Health (JBI) 2025 (30) | PubMed, Embase and PsycInfo | 2025 Jun | 44 |

Subsequent to creating the repository of studies above, we sourced studies included in two other recent, wide-ranging evidence reviews (26, 27). Both the Utah review (report for the Utah Department of Health and Human Services) and the RAND review covered medical and other interventions for young people with gender dysphoria (i.e. not solely puberty suppression). Studies in these two reviews (not already identified from the reviews in **Table 2**) were screened along with the records retrieved from the database searches (see below).

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License


FOR PUBLIC CONSULTATION

### 3.2.2 Database searches

We searched MEDLINE, Embase, PsycInfo (all via Ovid), CINAHL Complete (EBSCO) and Europe PMC on 6 March 2026. The search of Europe PMC was limited to preprints. Searches were not limited by language, status or format of publication. Since searches for the most recently published reviews listed in the table above were conducted in 2024 and 2025, we limited searches from 2024 onwards.

The search strategy for MEDLINE combined MeSH and free-text terms for the population (gender incongruence/dysphoria) and intervention (puberty suppression). The selection of search terms was informed by consulting the search strategies of several of the reviews listed in the table. Once the MEDLINE search strategy was finalised, it was translated for the other databases, incorporating database-specific thesaurus terms, as appropriate (see Supplement, Appendix S1.A).

Concurrent to running the database searches, the MEDLINE search strategy was independently reviewed using the PRESS checklist (50) by an information specialist familiar with the topic. The revised search strategy, incorporating suggested changes, retrieved 21 additional records in MEDLINE. Since none of these was relevant for this review, the other database search strategies were not revised.

The performance (relative recall) of the MEDLINE search was evaluated against the 97 MEDLINE-indexed records included in the systematic reviews and evidence reports listed in the table above. The search identified 84 of these studies. None of the 13 studies missed by the search was eligible for this review – most either evaluated the effects of gender affirming hormone therapy (not puberty suppression) or investigated characteristics of transgender youth (with no intervention reported).

### 3.2.3 Register searches

We searched the following registers (not restricted by date) on 31 March 2026 for details of any ongoing, completed or unpublished studies. Search terms are provided in the Supplement, Appendix S1.A. Searching for completed or unpublished studies in registers provides an additional check for potentially eligible studies and permits consideration in assessing missing results.

- WHO International Clinical Trials Registry Platform (trialsearch.who.int)
- ClinicalTrials.gov (clinicaltrials.gov)
- ISRCTN (www.isrctn.com)

### 3.2.4 Searching other sources

For each study included in the review, we manually checked the list of references. We also checked with NHMRC and the Committee for details of any studies they were aware of, but which were not retrieved by our search.

Since eligible studies had to be published, peer-reviewed reports (with the exception of preprints), we did not search grey literature sources. Studies published as conference abstracts were linked to published reports where possible or tagged as awaiting classification if no published full report was found. As a final step, for studies included in the review, we searched PubMed in May 2026 to check whether any were listed as retracted or had

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

expressions of concern, errata or corrections associated with them. One study (Turban 2020) corrected a result in one of the tables, but this had no bearing on the data we extracted for the review.

## 3.3 Data collection and analysis

### 3.3.1 Selection of studies

The pool of included studies identified from existing systematic reviews and evidence reports were imported into Covidence for full-text screening. Studies that did not meet the eligibility criteria were excluded and the reason for exclusion recorded in Covidence and Excel (see 'Studies assessed' tab Supplement 2). Records listed as excluded from the reviews by Miroshnychenko and Tornese were imported to Covidence for title/abstract screening.

Records from the database and register searches were imported into EndNote and duplicates removed. The remaining records were added to Covidence for title/abstract screening. Records deemed potentially eligible were then screened at full text.

Records at both title/abstract and full text were independently screened by two reviewers (SB and SM). At full-text review, one reviewer (SB) recorded the reasons for exclusion against all criteria. Disagreements at either stage of screening were resolved by consensus among members of the review team.

Studies that met the eligibility criteria, but for which results were not available in a published report, are listed as 'Ongoing studies'. Studies published as abstracts for which no full report is available are listed as 'Studies awaiting assessment'.

#### Dealing with duplicate and companion publications

Multiple publications to the same study (e.g. registry entries, protocols, abstracts, etc.) were identified at the data extraction stage and linked to the same study identifier.

### 3.3.2 Data extraction and management

Study data were collected and managed using REDCap electronic data capture tools hosted at Monash University (51, 52). Three authors pre-tested the quantitative data extraction and coding form on three studies (sufficient to achieve consistent coding; see 'Study results' for details). These studies were purposefully selected from the included studies to cover the diversity of data types and design characteristics anticipated in the review. One author reviewed the extracted and coded pilot data for completeness, accuracy and consistency, with discussion among all authors. Revisions to the data extraction form and guidance (embedded within the form) were made as required to maximise the quality and consistency of data collection.

Data requested by the ONHMRC is either listed below, listed under 'Types of outcomes', or addressed in the synthesis.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

### Characteristics of included studies

For each included study, one review author (SB) extracted study characteristics using a pre-tested data extraction and coding form. Uncertainties were resolved through discussion with a second author (SM), and advice sought from a third reviewer, the senior biostatistician (JM) for more complex scenarios as needed. Advice from the Committee was only required in one instance (related to outcome categorisation). The categorisation of study design and other design features was verified during the risk of bias assessment by another author (MP, JM during pilot of risk of bias [RoB]).

We extracted the following study characteristics:

1. Study identifiers and characteristics of the study design
    - Study references (multiple publications arising from the same study were matched to an index reference; coded as index paper, protocol, registry entry, results paper 1, 2, …)
    - Study name, location, commencement date, and other study identifiers
    - Study design (categorised as non-randomised study of intervention ['non-randomised controlled trial', 'prospective cohort study', 'retrospective cohort study', 'target trial emulation analysis', 'case-control study', 'controlled interrupted time series study', 'controlled before-after study', 'uncontrolled before-after study', 'cross-sectional study'])
    - Any additional design features needed to interpret the study and assess risk of bias
    - Funding sources and funder involvement in study
    - Financial and non-financial interests declared by investigators
    - Other declared information about the investigator team that may be pertinent to their expertise in research involving people who are trans and gender diverse, including the involvement and role of people with lived experience
2. Characteristics of each intervention group (including comparator groups)
    - Characteristics of the intervention reported by domains of the Template for Intervention Description and Replication (TIDIER) checklist (53), including setting of care, health professionals involved in delivering care, type of puberty suppressant, dose, route of administration, duration of treatment (see also point 3, Characteristics of participants).
    - Details of any co-interventions (supporting or concurrent gender affirming care)
    - Number of participants: in each group, at follow-up for the selected outcome, and included in analysis and reasons for loss to follow-up
    - Details of the methods used to collect information on the intervention allocated or received by each participant (i.e. information used to classify intervention status)
    - Any criteria used to determine eligibility for puberty suppressants (e.g. diagnostic, Tanner stage, etc.) whether for study eligibility or usual criteria used in practice
3. Characteristics of participants
    - Participant eligibility criteria (verbatim or precis)
    - Age (mean or median, range; separated by gender identity if reported)
    - Sex assigned at birth (% assigned female at birth, % assigned male at birth, % other, or not reported)
    - Gender identity (% female, % male, % non-binary, % unsure, % prefer not to say or not reported)

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

- Age and pubertal stage at which puberty suppression was commenced and discontinued (irrespective of reason)
- Proportion who proceed to other gender affirming care
- Equity related characteristics, in particular diversity of participants considering rurality, socioeconomic status, neurodivergence, disability, culturally and linguistically diverse groups (PROGRESS+) (54)
- Other characteristics of importance within the context of each study

4. Outcomes assessed
    - Outcomes measured (a list of all outcomes, irrespective of eligibility for this review, categorised according to the domains specified for the review, or as 'other' if none of the outcome domains apply)
    - See 'Study results' for data pertaining to the outcomes selected for inclusion in the summary and synthesis of results.

In the draft protocol, we reported that we would consider using data from characteristics of studies tables in Taylor et al. (32) and Tornese et al. (34) to expedite data collection if these data can be verified from the study reports. However, we revised this decision ahead of commencing the review because these reviews did not provide complete characteristics needed for our review and the process of extracting and summarising study characteristics enhanced our understanding and interpretation.

### Study results

Three statisticians pre-tested the results section of the data extraction form on three studies (SLT, DC, JM). The remaining results were extracted by two statisticians (SLT, DC), who sought advice from the senior biostatistician (JM) as required. For outcomes selected for inclusion in the summary and synthesis of results, we extracted:

- Outcome domain: categorised according to the domains specified in the outcomes section (3.1.4) (e.g. quality of life, body image)
- Outcome as described in the included study (verbatim or precis)
- Measurement method (e.g. Utrecht Gender Dysphoria Scale), information required to interpret measures (scale range and direction, minimally important difference) and time point (exact, and timeframe categorised as 'immediate' or 'longest follow-up')
- Results including: summary statistics within group (mean change, standard deviation of change, sample size, and any other reported statistics [e.g. correlation between before-after observations, confidence interval for mean change, test-statistics and p-values]), summary statistics by group (means and standard deviations, or number of events for dichotomous outcomes, and sample size for both), estimates of intervention effect (e.g. mean differences, adjusted mean differences, adjusted odds ratios), standard errors of the intervention effect, confidence intervals for the intervention effect, and related test-statistics (e.g. t-values) and p-values. For adjusted estimates, we extracted information on the analysis method, how confounding was adjusted, and which confounders were adjusted for.
- Data required to support risk of bias judgements (see Section 3.3.3)

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

### 3.3.3 Assessment of risk of bias

We identified only non-randomised studies and therefore describe here only our methods for assessing RoB in these designs. Our protocol also includes description of our planned methods for assessing RoB in randomised trials, had we identified any. We used version 2 the ROBINS-I tool for non-randomised studies of interventions (43, 55, 56). The assessment was made for each critical outcome included in the synthesis. Our assessment was based on the effect of assignment to the intervention, which in non-randomised studies is commencement of puberty suppression. Version 2 of ROBINS-I addresses the following six domains:

1. bias due to confounding
2. bias in classification of interventions
3. bias in selection of participants into the study
4. bias due to missing data
5. bias in measurement of outcomes
6. bias in the selection of the reported result

In our overall judgement for domains 1 and 3, we considered whether opposing residual confounding (i.e. the direction of bias) or evidence of a dose response mitigated concerns about bias due to confounding or selection of participants into the study (56, 57). The potential for large effects to offset any concerns about bias were considered when assessing the certainty of the synthesis result (see 3.3.10 Summary of findings tables and assessment of certainty of the evidence).

To promote concordance, the assessment was piloted by two review authors (including a statistician (JM)) on 3 studies, across two different designs. Three authors (MJP, JM, SB) discussed these assessments to reach a shared understanding of the design features specific to the topic that may introduce bias. In addition, we developed a template to facilitate consistent reporting of the risk of bias judgements. The template provided standardised wording for describing factors / actions that may introduce bias in each domain, supplemented with study specific details. One review author (MJP) then applied the ROBINS-I tool (42, 56) to the selected results from each study, and the assessments were verified by a second author (SB) who independently extracted the study design features and had a detailed understanding of the limitations of each study. Disagreement between review authors was resolved through discussion, and a third review author adjudicated where agreement was not reached. We derived an overall summary of the risk of bias from each assessment, following the algorithm in the ROBINS-I guidance (42, 56) (see Table 3).

**Table 3.** Algorithm used to derive an overall summary of the risk of bias for each result

| Judgement | How reached |
|---|---|
| **Low risk of bias except for concerns about uncontrolled confounding** | Low risk of bias except for concerns about uncontrolled confounding in Domain 1 and Low risk of bias in all other domains. |
| **Moderate risk of bias** | At least one domain is at Moderate risk of bias, but no domains are at Serious risk of bias or Critical risk of bias. |

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

| | |
|---|---|
| **Serious risk of bias** | At least one domain is at Serious risk of bias, but no domains are at Critical risk of bias. |
| **Critical risk of bias** | At least one domain is at Critical risk of bias. |

If we judged a result to be at "critical" risk of bias on the first domain (bias due to confounding), we did not assess other domains, since the overall risk of bias for the result would be "critical" by default.

### Selection of effects for RoB assessment and synthesis

When multiple effects of the intervention using different approaches were presented in a study report, we selected one effect for inclusion in the meta-analysis and for RoB assessments. For NRSIs, the selected effect was the estimate adjusted for critical confounding domains (see next Section), if available. When this effect was not available, we selected (or computed) the unadjusted effect estimate. Our protocol includes description of the hierarchy that would have been used to select the effect estimate for randomised trials, had we identified any.

### Pre-specification of confounding factors and co-exposures

Confounding domains are "prognostic variables (factors that predict the outcome of interest)" that also predict whether an individual receives one or the other interventions of interest (56). ROBINS-I defines important confounding domains as those "for which, in the context of [a specific] study, adjustment is expected to lead to a clinically important change in the estimated effect of the [exposure]". We identified the following confounding domains as *critical* to adjust for through design or analysis:

- Age at initiation of puberty suppression (used as a proxy for pubertal stage) OR Pubertal stage at initiation of puberty suppression
- Sex assigned at birth (AFAB; AMAB)
- Social factors, which could include one or more of:
    - Rurality (e.g. metropolitan, regional, rural, remote)
    - Socio-economic status
    - Supportive parent or caregiver or others (yes/no)
    - Gender minority stress (e.g. yes/no)
    - Ethnicity / race
- For an independent control group, if available, baseline measure of the outcome (when the outcome can be measured at baseline).

In making the assessment we considered the relevance of the confounding domains to the particular outcome; for example, social factors are unlikely to explain the halt in puberty, so adjustment would not be needed. Confounding domains were identified prior to assessment of studies through targeted consultation with content experts on the Committee.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

### 3.3.4 Measures of treatment effect

The effect measures selected or computed to quantify the effects of puberty suppressants depended on the study design, outcome type, and the number of studies contributing to a meta-analysis.

For continuous patient reported outcomes (gender dysphoria, body image and satisfaction, social functioning, mental health – psychological functioning), we used the standardised mean difference, using the Hedges' (adjusted) *g*. We standardised the effects for these outcomes because different measurement instruments were sometimes used across studies within a meta-analysis to measure the same underlying construct, and to facilitate consistent interpretation across meta-analyses. For uncontrolled before-after designs, and where estimates came from regression models with an independent comparison population, we standardised the calculated effects using the standard deviation of the observed outcome values, rather than a standard deviation that incorporated the correlation between observations. For continuous physical outcomes (penile length, SAE – QT interval prolongation), we used the mean difference for studies with an independent comparison population and the mean change from baseline for uncontrolled before-after studies. We did not standardise the effects for these outcomes because the scales are generally readily interpretable (e.g. cm).

For count outcomes (SAE – idiopathic intracranial hypertension), we calculated rates. For time-to-event outcomes (mental health – suicidality), we reported the hazard ratio (directly from the study). Finally, for dichotomous outcomes from studies with an independent comparison population (breast cup size B or larger), we had planned to calculate an odds ratio, because empirical evidence suggests that relative effect measures tend to be more consistent than absolute measures (58), and then convert this to a risk difference. However, as only one study contributed to the meta-analysis, we directly calculated the risk difference. For dichotomous outcomes from uncontrolled before-after studies (menstruation), we had planned to calculate the within-person odds of change. However, most studies reported the number of youth menstruating at follow-up only for the subset of youth who were menstruating at baseline, precluding calculation of the within-person odds of change. Instead, we calculated conditional (on menstruating at baseline) proportions using the Freeman-Tukey double-arcsine variance-stabilising transformation, which does not require a continuity correction for zero cells.

#### Thresholds for interpreting the effects

Following the GRADE approach for interpreting findings, all interpretations were based on where the combined estimate of effect (point estimate) lies in relation to a prespecified threshold for an important effect (i.e. whether an important effect is observed or not) and the direction of effect (beneficial or harmful) (59). Thresholds for interpreting the size of effect were set in absolute terms (e.g. for continuous outcomes an SMD or MD, and for dichotomous outcomes the absolute risk difference).

Given the diversity of outcome measurement methods that were anticipated in this review, it was not feasible to prespecify thresholds. Instead, we proposed thresholds for each outcome to the Committee for discussion and endorsement ahead of presenting any results. Where results from different outcome measures were combined using an SMD, we used a common version of Cohen's guiding rules for interpretation (an SMD of 0 to <0.2 represents a trivial effect, 0.2 represents a small effect, 0.5 a moderate effect, and 0.8 a large effect) (60). The interpretation was initially based on whether there was an important effect or not (59, 61, 62), for which the small effect threshold (SMD of 0.2) was used as the threshold for an important difference. To illustrate the approach, if

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

the SMD fell within the prespecified range of -0.2 to 0.2 (i.e. within both thresholds), the effect of puberty suppressants was considered to be no different from control. An SMD above 0.2 or below -0.2 was interpreted as an important effect. The thresholds are reported in the footer to the Summary of Findings table in the results.

### 3.3.5 Unit of analysis issues

In this review, we anticipated unit-of-analysis issues may have arisen from two types of studies: (i) studies that used individual-level matching between those who did and did not receive puberty suppressants based on particular characteristics (e.g. age), and did not account for this matching in the analysis; and (ii) uncontrolled before-after studies in which the same individuals provided measurements before and after puberty suppressants, but these repeated measurements were analysed as if they were independent. For (i), we identified no studies that used individual-level matching (except for ineligible comparator groups). For (ii), we undertook a paired analysis, assuming a correlation coefficient of 0.5 between before and after measurements for our primary analysis, based on typical within-person correlations observed in treatment arms of randomised trials (63). We examined the sensitivity of this assumption using a range of correlation coefficients (see Sensitivity analyses).

### 3.3.6 Dealing with missing data

We did not contact study authors to obtain missing information (e.g. study characteristics, description of conduct of the study) or aggregate level statistics (e.g. missing standard deviations). We attempted to calculate statistics necessary for meta-analysis using algebraic manipulation of reported statistics (e.g. computing the standard deviation from the confidence interval of the mean). Details of our algebraic manipulations and imputations are available in Supplement 1, Appendix S1.C.

Our approach for dealing with missing outcome data within the studies was through sensitivity analyses, where studies judged to be at critical risk of bias were excluded (Section 3.3.9 Data synthesis). Risk of bias ‘due to missing outcome data’ is considered within the overall bias judgement for a study.

### 3.3.7 Assessment of heterogeneity

We assessed statistical heterogeneity of the intervention effects visually by inspecting the overlap of confidence intervals on the forest plots, formally testing for heterogeneity using the $\chi^2$ test (using a significance level of $\alpha$=0.1), and quantifying inconsistency using the $I^2$ statistic (64).

### 3.3.8 Assessment of biases due to missing results and integrity

We planned to use a framework for assessing risk of bias due to missing results in which an assessment is made for each meta-analysis regarding the risk and potential impact of missing results from studies (termed ‘known-unknowns’) and the risk of missing studies (termed ‘unknown-unknowns’) (65). However, we were unable to apply this framework. As is typical of non-randomised studies, only one of the included studies had a registry entry or a protocol specific to the reported analyses (66) (although several protocols were identified for multi-centre collaborations, from which studies are now being published). In addition, there were too few studies contributing to the meta-analyses to generate informative contour enhanced funnel plots to examine whether there was evidence of small-study effects (67). Instead, we made a judgement of whether there was ‘undetected’ or

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

'suspected' reporting bias for each of the comparisons in our GRADE assessment (Section 3.3.10 Summary of findings tables and assessment of certainty of the body of evidence). Given the large number of ineligible studies identified through the screening process as addressing similar questions, we judged that there was no evidence to suspect reporting bias.

While there is increasing interest in assessing the integrity of randomised trials, the INSPECT-SR tool is yet to be adapted for non-randomised studies (68). In the absence of a well-validated tool to assess integrity, and the low likelihood of finding information required for assessment, we did not perform formal integrity checks.

## 3.3.9 Data synthesis

### Meta-analysis

Prior to undertaking any meta-analysis, we carefully considered the clinical and methodological diversity of the included studies, including variation in populations, interventions, study designs, effect measures, and risk of bias, to determine whether meta-analysis was appropriate. In any meta-analyses we undertook, we used inverse-variance random-effects meta-analysis, since we anticipated there could be statistical heterogeneity arising from the likely aforementioned differences across studies. These analyses used the restricted maximum likelihood estimator (REML) of between-trial heterogeneity variance and the Hartung-Knapp-Sidik-Jonkman confidence interval method. An exception to this, and in line with advice from the Cochrane Handbook (69), was when the estimated heterogeneity variance was zero, where we used a Wald-type confidence interval method with standard normal quantiles. A further exception was for rare events or where there was only one study that provided two estimates, for which we used an inverse-variance fixed-effect model instead of a random-effects model. For rare events, estimating the heterogeneity variance can be difficult. Where a single study provided two estimates, a fixed-effect model was appropriate because the estimates represented different strata within the same study. Forest plots were used to visually depict the intervention effect estimates and their 95% confidence intervals.

### Summary and synthesis when meta-analysis is not possible

We had prespecified methods for reporting results and for considering alternative synthesis methods if meta-analysis could not be undertaken. However, for all outcomes, we were able to undertake meta-analysis or, where only an effect estimate from an individual study was available, we reported the effect estimate and its 95% confidence interval.

### Subgroup analysis and investigation of heterogeneity

We had planned, if feasible to undertake a subgroup analysis to examine whether age (a proxy for stage of pubertal development) and sex assigned at birth explained any observed statistical heterogeneity in the intervention effects for relevant outcomes. In addition, where possible we had planned to stratify by different measures for an outcome domain to investigate if this explained any observed statistical heterogeneity in the intervention effects. Given that all meta-analyses included fewer than five studies and showed little or no heterogeneity, we did not undertake these analyses. An exception was for outcomes penile length, breast cup size, and breast resection volume, for which effects estimates were reported in only one study (70), and separately by Tanner stage (2/3 and 4/5). We therefore maintained this stratification in the review.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

### Sensitivity analyses

We undertook the following sensitivity analyses examining if the meta-analysis estimates were robust to the:

- inclusion of studies judged to be at critical risk of bias; and,
- choice of the imputed correlation coefficient in uncontrolled before-after studies in which we undertake a re-analysis. We examined the impact of using a range of correlation coefficients (0, 0.3 and 0.7).

We had planned to also undertake sensitivity analyses to examine the impact of fitting a fixed-effect model as an approach to investigating potential small-study effects, and to exclude studies that had not adjusted for matching. We did not undertake these analyses because there was little or no heterogeneity in any of the meta-analyses, meaning that the random-effects estimates and confidence intervals would not differ from those obtained using a fixed-effect model. In addition, no studies used individual-level matching.

Results of the sensitivity analyses were tabulated, including the meta-analysis estimate (and its confidence interval), along with details of the original and sensitivity analysis assumptions (see Supplement 4).

## 3.3.10 Summary of findings tables and assessment of certainty of the evidence

We prepared a GRADE summary of findings table, reporting results for all critical and important outcome domains. For each result, one author (SB) applied the GRADE approach to assess our confidence in where the effect lies relative to our threshold for an important effect (the certainty of evidence) (see 3.3.4 Measures of treatment effect). A second author independently verified the assessment (JM) without knowledge of the first assessor's judgements. In accordance with detailed GRADE guidance (57, 71-73), an overall GRADE of high, moderate, low or very low certainty is reported for each result based on whether there are serious, very serious, extremely serious or no concerns in relation to each of the following domains.

1. **Risk of bias**. We assessed the overall risk of bias across all studies contributing to each synthesised result, considering the weight studies rated at high risk of bias contribute to the analysis. Serious, very serious or extremely serious concerns were more likely if studies at higher risk of bias contributed considerable weight in the analysis and sensitivity analyses indicate that removing these studies changed the size of the effect importantly (i.e. leading to a different interpretation in relation to the specified thresholds) (see 3.3.9 Sensitivity analyses) (57, 72, 74).
2. **Inconsistency**. We assessed whether there was important, unexplained inconsistency in results across studies considering the overlap of confidence intervals (non-overlap indicating potentially important differences in direction or size of effect), statistical measures that quantify and test for heterogeneity ($I^2$ statistic, $\chi^2$ test), results of subgroup analyses (see 3.3.7 Assessment of heterogeneity), and where the point estimates lie in relation to the threshold for an important effect (if all to one side of a threshold, we were less concerned) (75, 76). Where a result is based on a single study, inconsistency was not rated down as per GRADE guidance.
3. **Imprecision**. We assessed whether the confidence interval for each pooled effect estimate crossed our threshold for an important effect (meaning that the confidence interval is compatible with different interpretations, for example an important effect and little or no difference in treatments). In judging imprecision, we used the thresholds for an important effect as described in 3.3.4 Measures of treatment effect (also reported in the Summary of findings table footer). We rated down for serious imprecision when

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

the confidence interval crosses one threshold, very serious when two thresholds are crossed, and extremely serious when the confidence interval is so wide that the estimate is considered uninterpretable (59, 62, 77). Where the extent to which the confidence interval crossed the threshold was modest (e.g. 2nd decimal place for an SMD), and both the upper and lower bound of the confidence interval indicated the same direction of effect, we did not rate down to avoid different certainty judgements for results that are not materially different.

4. **Indirectness**. We assessed whether there were important differences between the characteristics of studies included in each synthesis and the review question, such that the effects observed may not apply to the review question (i.e. the applicability of the evidence) (78). Empirical evidence is required to justify a concern that these differences are likely to influence the size of effect.
5. **Dissemination bias**. Our judgement of suspected dissemination bias was based on an informal assessment of bias due to missing results, due to the absence of registry entries and protocols for observational studies (see 3.3.8) (79).
6. **Upgrading domains** (large effect size, dose response gradient, opposing plausible residual confounding). We considered upgrading for large effects (informed by advice from the Committee on thresholds for large effects on the pubertal development and breast tissue resection outcomes). Where the effects are large or very large the evidence may be rated up, with rating up more likely if both ends of the 95% confidence interval indicate a large effect, effects are consistent, and the previous trajectory of the outcome is reversed. There was no dose related data to assess dose response. Opposing residual confounding was considered as part of the ROBINS-I assessment (judging whether this mitigated concerns that observed effects are due to residual confounding), which we accounted for in the overall risk of bias for each study (57, 72).

Using GRADE decision rules, we derived an overall GRADE for the certainty of evidence for each result included in the summary of findings table (57, 72, 80). In line with GRADE guidance for when ROBINS-I is used, a result from a body of evidence comprised of NRSIs begins as 'high' certainty evidence (score=4),and can be rated down (-1 or -2) for serious or very serious concerns on any GRADE domain that reduces confidence in the effect of puberty suppressants (as determined by the prespecified thresholds) (57, 72, 80). The option to rate down by -3 may be considered if there are extremely serious concerns about risk of bias or imprecision (where results are compatible with both important benefit and important harm) (62, 81). The evidence can be rated up for large (+1) or very large (+2) effects.

Summary of findings tables were prepared using the GRADEpro GDT software (www.gradepro.org). The tables included:

- estimates of the effects of puberty suppressants reported as SMD or MD for continuous outcomes (absolute effects), and for binary outcomes relative and absolute effects
- the overall GRADE (rating of certainty) and an explanation of the reason(s) for rating down (or up) (82)
- the study design(s), number of studies and number of participants contributing data
- a plain language statement interpreting the evidence for each comparison and outcome, following GRADE guidance for writing informative statements (83).

We did not have any outcomes with evidence from more than one study design (e.g. before-after and cohort), so there was no need to present multiple results for the same outcome (84).

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License


FOR PUBLIC CONSULTATION

### 3.3.11 Interpretation of findings (evidence statements)

When interpreting results, we followed GRADE guidance for writing informative statements (83). All interpretations were based on where the point estimate lies in relation to the prespecified thresholds for an important effect (an important effect or not) and the direction of effect (beneficial or harmful). When we had evidence suggesting a large effect (sufficient to rate up certainty), we confirmed this interpretation with the Committee. The certainty of evidence is communicated by qualifying the interpretation of effect (e.g. 'probably' improves for moderate certainty; 'likely' improves is an alternative). For low certainty evidence the interpretation is qualified with the word 'may'. For example, 'puberty suppressants may improve [outcome]' indicates that the point estimate lies above the threshold for important benefit (an SMD >0.2) and that the evidence is of low certainty. For very low certainty evidence, we provide an interpretation of the result except in certain circumstances (extreme imprecision or inconsistency), which is consistent with the usual GRADE approach for non-randomised evidence and one of two options GRADE provides for interpreting findings based on very low certainty of evidence (83).

# 4 Results

## 4.1 Search results

The database searches retrieved 830 records. After removing 270 duplicates in EndNote, 560 records were imported into Covidence and deduplicated against the 100 studies already added to full-text review from the existing systematic reviews and evidence reports (see table 2). A further 15 duplicate records were removed, leaving 545 records from the database searches to screen at title/abstract. Forty-five records were retrieved from the three register searches, of which 42 were imported to Covidence for screening (after removing duplicates).

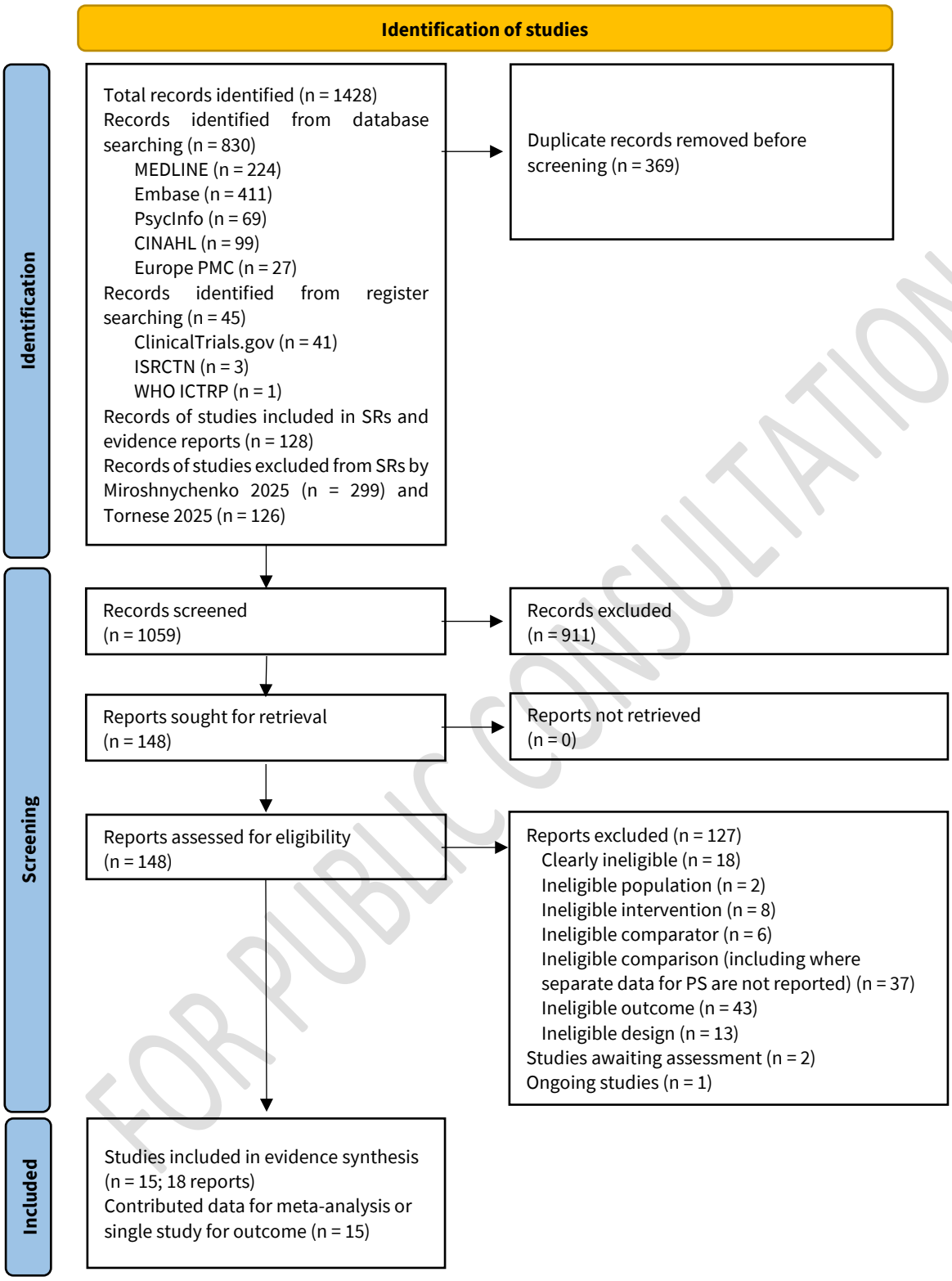
In a separate Covidence folder we imported the records excluded at full-text review from the systematic reviews by Miroshnychenko (n=299) and Tornese (n=126), after first deduplicating against the 100 studies identified from the existing systematic reviews and evidence reports. Eighty-one duplicates were removed, leaving 344 to be screened. To these we added 28 unique records from the Utah and RAND reviews. Of the 372 records screened, 365 were excluded at title/abstract, six were excluded at full-text review, and one was a conference abstract linked to an included study (Turban 2020).

In total, 15 studies (18 reports) were included in the review (70, 85-99). Thirteen studies (16 reports) were sourced from among the 100 studies included in existing SRs and evidence reports, and two studies (two reports) were included from the update database searches (Lahaije 2025 and Zaitoon 2025). We identified one ongoing study (PATHWAYS trial) and two abstracts as awaiting assessment (Sun 2025a and Sun 2025b) from the update search.

The flow of reports and studies through the review is shown in the PRISMA flow diagram (**Figure 2**). Details of studies excluded at full-text review are given in the Table of Excluded Studies (Supplement 2). No additional studies were identified by the Committee.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License




FOR PUBLIC CONSULTATION

**Fig 2** | PRISMA diagram showing the flow of studies through the review.

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

## 4.2 Description of studies

A summary of the characteristics of all studies contributing to this review follows. The characteristics of the subset of studies contributing to each result are synthesised alongside the effects. Characteristics of individual studies are tabulated in Supplement 2 (an Excel spreadsheet filterable by outcome or other characteristics). Risk of bias assessments are reported in Supplement 3.

### 4.2.1 Study setting and design

*Location and setting.* Eight of the 15 included studies were in continental Europe, 3 studies were in the United Kingdom, 3 studies were in the United States, and one study was in Israel.

- Netherlands: 5 studies from the Vrije University Medical Centre Gender Clinic in Amsterdam
- UK: 3 studies at the National Health Service Gender Identity Service at the Tavistock Centre
- US: one national survey, one study in a paediatric hospital gender clinic and one study that analysed data from 5 paediatric hospitals
- Germany: 1 study at the Hamburg Gender Identity Service
- Israel: 1 study in the Israeli Children and Adolescents Gender Clinic, located at Dana-Dwek Children's Hospital, Tel Aviv
- Italy: 1 study at the Florence Gender Clinic
- Sweden: 1 population-level study using data from national databases

There were multiple reports from two gender clinics (Amsterdam and UK), some of which had data that were likely to have come from the same participants. Where this was the case, we included the data if the outcomes reported addressed different outcome domains. If no unique outcome data were reported, the paper was listed as a secondary report of a study (see list of included studies). In one instance, we could not rule out the possibility that data from some participants were included in both papers, so we did a sensitivity analysis to ensure that including only one of the papers did not alter interpretation of findings.

*Study designs.* We independently coded the design of studies based on design features prespecified in the protocol. Using this classification, one included study was a prospective cohort (Costa 2015 (85)), 7 were retrospective cohort studies (Becker Hebly 2021(100), Karamanis 2023 (89), McGregor 2024 (92), Nunes-Moreno 2025 (94), Turban 2020 (96, 97), van de Grift 2020 (70), van der Miesen 2020 (98)) and the other 7 studies used an uncontrolled before-after design (Carmichael 2021 (66, 93), de Vries 2011 (86, 87), Fisher 2024 (88), Lahaije 2025 (90), Lavender 2023 (91), Schagen 2016 (95), Zaitoon 2025 (99)).

While study design in and of itself does not convey whether the study is at risk of bias, in general, uncontrolled before-after studies are at critical risk of bias for most outcomes. This is because it is usually not possible to know whether any change is due to the intervention, other interventions to which an individual is exposed (such as gender affirming care), or change over time (such as temporal changes in mental health). We consider pubertal development to be an exception because the typical trajectory of pubertal development is predictable for the vast majority of adolescents (e.g. in terms of development of physical characteristics), so effects that are in the opposite direction to that expected are likely to be attributable to puberty suppression, especially if the effects are large.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

### 4.2.2 Characteristics of those receiving puberty suppressants

We included studies in which all (or a large majority) of participants were likely to have received puberty suppressants before 18 years of age. We did not limit by age at follow-up (a decision taken ahead of screening studies based on feedback on the protocol from the Committee and Reference Panel members). None of the studies reported limiting their sample by gender or sex assigned at birth. The majority of studies reported the number of participants by sex assigned at birth (7232 assigned male at birth [AMAB]; 5179 assigned female at birth [AFAB]). When gender was reported, it was generally unclear whether this was self-identified by participants or assumed by the investigators. Seven studies reported separate results data by sex assigned at birth or by gender (see Supplement 2, Characteristics of included studies table for details of available data).

We sought data on equity related characteristics (race/ethnicity, rurality, socio-economic status, neurodivergence, disability, other factors). There was limited reporting of equity related data, and the available data were sometimes reported at the level of cohort. In such instances, data were not available for each treatment group and sometimes included data for comparator groups that were ineligible for this review (e.g. those who had received GAH; see Supplement 2, Characteristics of included studies). Five studies reported on race/ethnicity (three in the US, one in the Netherlands, one in the UK). Two studies reported on socio-economic status, one using an established measure (Winkler index). Three studies collected data about neurodivergence, two of these reported data. No studies reported on rurality or disability. Seven studies reported on parents' level of education.

### 4.2.3 Puberty suppression interventions

#### Assessments prior to prescribing puberty suppressants

Studies based in gender clinics followed a comprehensive interdisciplinary assessment process to determine eligibility (or 'readiness') for puberty suppressants (reported either in the included study or a secondary paper, the latter including descriptions of standard gender affirming care delivered in the clinic). Most cited the Endocrine Society guidelines as the basis for assessment, with additional considerations for gender affirming care informed by WPATH guidelines.

Commonly reported criteria included:

- persistent gender dysphoria (considering pervasiveness and intensity, with variable reporting of the use of formal diagnostic criteria)
- first physical changes of puberty (Tanner 2 or above)
- an absence of serious psychosocial comorbidity that would interfere with assessment or treatment (e.g. repeated episodes of psychosis)
- no medical contraindications
- a request for gender affirming intervention (usually, specifying that 'puberty suppression' was requested)

Other considerations included the adequacy of psychological and social support during treatment (e.g. from family or others), and "knowledge and understanding of both the physical and social consequences of gender affirmation" (for example, Becker-Hebly 2021, Costa 2015).

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

### Treatment protocols and duration

The primary aim of studies included in this review was to report evaluations of the effects of puberty suppressants, rather than detailed protocols for treatment. Nonetheless, some information can be ascertained on usual care delivered in gender clinics over the period in which the studies were conducted.

*Usual care*. Studies based in gender clinics reported regular follow-up (e.g. initially at 3 months, then 6 monthly) as a prerequisite for care, which typically included psychosocial support, medical review and care (e.g. by endocrinologist, paediatrician), routine blood testing to monitor metabolic and hormonal parameters (for example), and bone health checks (including scans and blood-based biomarkers). Some reported more specific elements of supportive care, for example fertility counselling and preservation, and coordination of care with mental health services for those with co-existing mental health conditions (e.g. Carmichael 2021, Becker-Hebly 2021).

*Type of GnRHa*. We limited inclusion to gonadotrophin-releasing hormone agonists (GnRHa), so effects of GnRHa were evaluated in all studies. All five studies that reported the type of GnRHa prescribed triptorelin administered via intramuscular injection (usually 3.75 mg injection every 28 days). Four of these studies were in Europe and one in Israel.

*Timing and duration of treatment*. The timing and duration of treatment with puberty suppression varied both within and across clinics (influenced by factors such as pubertal stage at first assessment, time from assessment to prescribing, and age at which GAH could be accessed). While it was possible to ascertain an approximate length of time on puberty suppressants at follow-up in most studies, the time from commencing to stopping puberty suppressants was rarely available. In part, this was because many studies had either a defined length of follow-up or a cutoff date for data collection that was before treatment cessation for some or all in the cohort. Of the four studies from which information was available, the typical duration of puberty suppressant treatment was 18 months to 2 years (de Vries 2011, Lavender 2023, McGregor 2024). The fourth reported “1 year to 4 or 5 years” (Carmichael 2021).

*Gender affirming hormones*. In general, we excluded studies where participants had also received gender affirming hormones (GAH), unless separate data were available that enabled effects of puberty suppressants to be isolated[6]. One study reported outcomes in adulthood that enabled comparison of the effects of puberty suppressants versus no puberty suppression on pubertal development and outcomes of future surgeries (specifically, serious adverse effects) (van de Grift 2020).

## 4.2.4 Comparators

*Care received in the comparator*. We included two types of comparators – one where there was an independent group of trans and gender diverse young people who did not receive puberty suppressants (retrospective and prospective cohort studies), and the other where those who received puberty suppressants served as their own controls (uncontrolled before-after studies). In the cohort studies, care received by the comparator group was

[6] The effects of gender affirming hormones will be examined in a future systematic review supporting the development of the Guidelines.

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License




FOR PUBLIC CONSULTATION

usually similar to that in the puberty suppressant group (i.e. gender affirming care or usual care, depending on the setting of the study). In uncontrolled before-after studies, outcomes are compared in the period before commencing puberty suppression to outcomes at follow-up. The care received in the period prior to suppression may be quite different, especially for those whose baseline data (used for the outcome measure taken in the 'before' period) was measured at first visit to a gender clinic.

### 4.2.5 Outcomes

*Outcomes*. A complete list of outcomes reported in each study (both eligible and ineligible) is provided in the characteristic of included studies (CoIS) table, together with length and frequency of follow-up. From the eligible outcomes (those prioritised for the guideline), we selected one outcome from each domain for inclusion in the meta-analysis (marked in the CoIS table with *).

Critical outcomes were eligible from studies of any design, whereas important outcomes were only included from studies that had an independent comparator group (prospective and retrospective cohorts). The length of follow-up varied; most studies reported results between 6 months and 2 years from commencing puberty suppressants.

## 4.3 Effects of puberty suppression

### 4.3.1 Summary of findings

Findings from the systematic review are presented in the GRADE summary of findings table (**Table 4**). The thresholds used to interpret each result are presented at the end of the table. Explanations for all GRADE judgements are in footnotes to the table. In the text following the SoF table, results are ordered by outcome importance. Outcomes prioritised by the Committee as critical for decision-making appear first (pubertal development, gender dysphoria, body image and satisfaction, and serious adverse events), followed by those prioritised as important but not critical. The rationale for prioritisation was reported a-priori in the protocol (see Methods, Types of outcomes; **Appendix 1a** outcomes table; **Figure 1** logic model).

**Table 4**. Summary of findings table

| **Puberty suppressants compared to no puberty suppressants for young people experiencing gender incongruence or dysphoria** |
|---|
| **Patient or population:** young people experiencing gender incongruence or dysphoria who commenced puberty suppression at <18 years of age |
| **Setting:** Any care setting was eligible. Care was provided in specialist gender clinics in 12 of 15 studies, paediatric hospitals (ED or inpatient admissions) in one study, and not reported in 2 studies (studies using national population data). Countries: 5 Netherlands, 3 UK, 3 USA, 1 Germany, 1 Israel, 1 Italy, 1 Sweden. |
| **Intervention:** puberty suppressants (gonadotropin-releasing hormone agonists, GnRH agonists). We included studies in which some participants had received gender affirming hormones (GAH) if effects of puberty suppressants could be isolated. We included results from one study where participants had also received GAH and gender-affirming surgery because the study examined potentially serious AEs for those who underwent future gender affirming surgery and compared outcomes for people who had received puberty suppressants with those who had not. |
| **Comparison:** no puberty suppressants. Gender affirming care (GAC) was provided to both groups in 12 of the 15 studies with results in this table. Three of the 15 studies did not report whether GAC was provided. In two of these studies, data were from a national population sample (one study from the USA that contributed data on mental health; one study from Sweden on idiopathic intracranial hypertension). The third was a multi-centre study using data from paediatric hospitals (a study in the USA on suicidality diagnoses). |

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

| Outcome<br>№ of participants<br>(studies) | Relative effect<br>(95% CI) | Anticipated absolute effects (95% CI) | | | Certainty | What happens |
|---|---|---|---|---|---|---|
| | | **no puberty suppressants** | **puberty suppressants** | **Difference** | | |
| **CRITICAL OUTCOMES** | | | | | | |
| Pubertal development (AMAB): penile length (cm)<br>follow-up: range 4 to 5 years<br>№ of participants: 103 (1 non-randomised study) | - | The mean penile length without puberty suppressants was **12.8** cm | - | MD **3.26 cm shorter** (3.44 shorter to 3.09 shorter) | ⊕⊕⊕◯<br>Moderate[a,b] | Puberty suppressants (GnRH agonists) likely suppress pubertal development importantly (shorter penile length) compared to no puberty suppression. Effects were larger among those who received puberty suppressants at Tanner stage 2/3 (MD 4.90 cm shorter, 95% CI: 4.63 to 5.17 cm shorter). [c] |
| Pubertal development (AFAB): menstruating<br>follow-up: range 3 to 12 months<br>№ of participants: 143 (4 non-randomised studies) | In all four studies, none of the young people who were at menarche before puberty suppressants (baseline: 15/19, 20/22, 49/65, 35/37) had menses after puberty suppressants (follow-up: 12, 3, 12 and 12 months respectively).<br>Proportion menstruating at follow-up: 0.00 (95% CI: 0.00 to 0.01) | | | | ⊕⊕⊕⊕<br>High[d] | Puberty suppressants (GnRH agonists) suppress pubertal development (menstruation) importantly compared to no puberty suppression. The effects were very large in all four studies. |
| Pubertal development (AFAB): breast cup size B or larger<br>follow-up: range 4 to 5 years<br>№ of participants: 126 (1 non-randomised study) | RD 47 fewer per 100 had a breast cup size of B or larger (95% CI: 32 fewer to 62 fewer) | | | | ⊕⊕◯◯<br>Low[a,e] | Puberty suppressants (GnRH agonists) may suppress breast growth importantly compared to no puberty suppression. Effects were larger among those who received puberty suppression at Tanner stage 2/3. |
| Gender dysphoria assessed with: Utrecht gender dysphoria scale (lower is better)<br>follow-up: range 12 to 24 months<br>№ of participants: 100 (3 non-randomised studies) | - | - | - | SMD **0.06 SD lower** (0.25 lower to 0.13 higher) | ⊕◯◯◯<br>Very low[f,g] | The evidence suggests that there may be little to no change in gender dysphoria between commencing puberty suppressants and follow-up, but the evidence is very uncertain.[h] |
| Body image and satisfaction assessed with: Body image scale (overall or secondary sex characteristics, lower is better)<br>follow-up: range 12 to 24 months<br>№ of participants: 118 (3 non-randomised studies) | - | - | - | SMD **0.15 SD higher** (0.03 lower to 0.33 higher) | ⊕◯◯◯<br>Very low[f,i] | The evidence suggests that there may be little to no change in body image and satisfaction between commencing puberty suppressants and follow-up, but the evidence is very uncertain.[h] |

FOR PUBLICATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

| Outcome<br>№ of participants<br>(studies) | Relative effect<br>(95% CI) | Anticipated absolute effects (95% CI) | | | Certainty | What happens |
|---|---|---|---|---|---|---|
| | | **no puberty suppressants** | **puberty suppressants** | **Difference** | | |
| SAE: idiopathic intracranial hypertension (IIH) assessed with: ICD10 code during inpatient visit or specialist outpatient care<br>№ of participants: 4230<br>(1 non-randomised study) | Incidence rate difference: 15.7 fewer IIH events per 100000 person-years among those receiving PS (95% CI: 37.4 fewer to 6.1 more per 100000 person-years) [j]<br><br>No cases of IIH were identified among individuals receiving GnRH agonists (410 people, median follow-up 1.4 years) | | | | ⨁⨁◯◯<br>Low[k,l] | Puberty suppressants may have little to no effect on the incidence of idiopathic intracranial hypertension (a serious but rare adverse outcome) compared to no puberty suppressants. |
| SAE: QTc interval assessed with: ECG<br>follow-up: range 2 to 3 months<br>№ of participants: 16<br>(1 non-randomised study) | Trans women: MD 20.5 ms longer (95% CI: 4.3 longer to 36.7 longer)<br><br>Trans men: MD 12.5 ms longer (95% CI: 8 shorter to 33 longer) | | | | ⨁◯◯◯<br>Very low[m,n] | Puberty suppressants may have little to no effect on QTc interval (which, if prolonged, is a risk factor for serious heart rhythm abnormalities) but the evidence is very uncertain. |
| Amount of breast tissue resected (grams) in future surgeries (SAE: invasive mastectomy)<br>follow-up: range 4 to 5 years<br>№ of participants: 174<br>(1 non-randomised study) | - | The mean amount of breast tissue resected in future surgeries (SAE: invasive mastectomy) without puberty suppressants was **684** grams | - | MD **248 grams lower**<br>(270 lower to 226 lower) | ⨁⨁⨁◯<br>Moderate[a] | For those undergoing future gender-affirming surgery, puberty suppressants probably result in less invasive mastectomy (i.e. a large reduction in the amount of breast tissue removed or resected). Effects were larger among those who received puberty suppressants at Tanner stage 2/3 (MD 540 grams less tissue resected, 95% CI: 605 lower to 475 lower). |
| **IMPORTANT OUTCOMES** | | | | | | |
| Satisfaction with care (gender-affirming prior to receipt of GAH) - not reported | - | - | - | - | - | |
| Quality of life and wellbeing assessed with: Kidscreen-27 or SF-8 (higher is better)<br>follow-up: range 18 to 24 months<br>№ of participants: 32<br>(1 non-randomised study) | - | - | - | SMD **0.77 SD higher**<br>(0.03 higher to 1.5 higher) | ⨁◯◯◯<br>Very low[o,p] | The evidence is very uncertain about the effect of puberty suppressants on quality of life and wellbeing (due to the extremely wide confidence interval). |

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

| Outcome<br>№ of participants<br>(studies) | Relative effect<br>(95% CI) | Anticipated absolute effects (95% CI) | | | Certainty | What happens |
|---|---|---|---|---|---|---|
| | | **no puberty suppressants** | **puberty suppressants** | **Difference** | | |
| Functioning (socially; family)<br>assessed with: Youth self-report (YSR - social problems scale); Child Global Assessment Scale (CGAS) (lower is better)<br>follow-up: range 18 to 24 months<br>№ of participants: 991<br>(4 non-randomised studies) | - | - | - | SMD **0.3 SD lower**<br>(0.45 lower to 0.15 lower) | ⨁⨁◯◯<br>Low[q,r] | Puberty suppressants may improve functioning (overall psychosocial, social) importantly. |
| Mental health - symptoms (psychological functioning)<br>assessed with: Youth self-report (overall, internalising) (lower is better)<br>follow-up: range 18 to 24 months<br>№ of participants: 4414<br>(4 non-randomised studies) | - | - | - | SMD **0.39 SD lower**<br>(0.68 lower to 0.11 lower) | ⨁⨁◯◯<br>Low[s,t] | Puberty suppressants may improve psychological functioning importantly. |
| Mental health - suicidality (attempts or thoughts, as diagnosed at ED or inpatient visit)<br>assessed from electronic health records at 5 paediatric hospitals<br>follow-up: mean 2 years<br>№ of participants: 3321<br>(1 non-randomised study) | **HR 0.79**<br>(0.47 to 1.31) | **Moderate risk**<br>16.0% | **13.0%**<br>(8 to 20) | **3.0% fewer**<br>(8.0% fewer to 4.0% more) | ⨁◯◯◯<br>Very low[u,v] | Puberty suppressants may reduce suicidality diagnoses but the evidence is very uncertain. Using the baseline risk for suicide attempts reported in Trans Pathways (an Australian national survey of young TGD people), puberty suppressants may lead to 3 fewer diagnoses of suicidality in an ED/inpatient setting per 100 people over 12 months. |
| Mental health - death by suicide - not reported | - | - | - | - | - | |

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

| Outcome<br>№ of participants<br>(studies) | Relative effect<br>(95% CI) | Anticipated absolute effects (95% CI) | | | Certainty | What happens |
|---|---|---|---|---|---|---|
| | | **no puberty suppressants** | **puberty suppressants** | **Difference** | | |
| AE: bone fracture - not reported | - | - | - | - | - | |
| AE: fertility - not reported | - | - | - | - | - | |

**The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
**CI:** confidence interval; **HR:** hazard ratio; **MD:** mean difference; **SMD:** standardised mean difference; **RD:** risk difference; **AMAB**: assigned male at birth; **AFAB**: assigned female at birth; **PS** = puberty suppression; **GnRH agonists** = gonadotropin-releasing hormone agonists, **GAC** = gender affirming care, **GAH** = gender affirming hormones, **GAS** = gender affirming surgery, **ED** = emergency department, **AE** = adverse event, **SAE** = serious adverse event, **cm** = centimetres, **ms** = milliseconds
**Thresholds.** The thresholds for an important effect (benefit or harm) were agreed prior to the GDC seeing results. These thresholds were used for interpreting point estimates and assessing certainty of the evidence and were as follows.

- Pubertal development – penile length: 10% difference in length (in centimetres). Shorter penile length indicates pubertal suppression.
- Pubertal development – menstruating: 10 fewer [or more] people menstruating per 100. Fewer people menstruating indicates pubertal suppression.
- Pubertal development – breast cup size B or above: 10 fewer [or more] people with a breast cup size of B of above per 100 people. Fewer people with a breast cup size of B or above indicates pubertal suppression.
- Gender dysphoria, body image and satisfaction, quality of life and wellbeing, functioning, mental health symptoms – psychological functioning: SMD of 0.2. A lower score (i.e. SMD of -0.2 or below) is better for all outcomes (i.e. less dysphoria, less body dissatisfaction, fewer problems with psychosocial functioning, fewer problems with psychological functioning) except quality of life (SMD of 0.2 or above is better).
- Serious adverse events – idiopathic intracranial hypertension (IIH): 20 more with incident IIH per 100000 person years. Fewer people with IIH is better. The threshold is set low because this is a rare adverse event (not an outcome that is a goal of treatment), Left undetected, IIH can have serious undesirable consequences.
- Serious adverse events – QTc interval: change from baseline of ≥30 ms. Based on ICH E14 / FDA guidance where an increase of ≥30 ms signals a potential concern ( >500 ms or of >60 ms over baseline are commonly used as thresholds for potential discontinuation)
- Serious adverse events – breast resection volume for those undergoing future gender-affirming surgeries: 10% difference in volume resected. A reduction in the amount of breast tissue removed or resected indicates less invasive surgery.
- Mental health symptoms – suicidality: 1 fewer [or more] person with a diagnosis of suicidality per 100 people over 12 months.

No eligible studies reported on satisfaction with care, mental health – death by suicide, adverse events – bone fracture or other functional measures of bone health, adverse events – fertility potential
**Control group risk**. The baseline risks for 'attempted suicide in the last 12 months' was from Trans-pathways 2017, a large Australian survey of trans and gender diverse young people. https://www.thekids.org.au/projects/past/trans-pathways/

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

## Explanations

a. Serious risk of bias (-1)/large effect (+1): study with 100% of weight in the analysis rated at moderate risk of bias. This effect is large (both the point estimate and either end of the 95% CI indicate large effects relative to the threshold, with the direction of effect reversed from usual pubertal development), offsetting concerns about risk of bias and justifying an upgrade for a large effect (+1).
b. Serious indirectness (-1): Tanner stage provides a more direct measure of pubertal development than penile length.
c. Compared to no puberty suppression, for those who commenced puberty suppression at Tanner stage 4/5 penile length was MD 2.10 cm shorter (95% CI: 1.87 to 2.33 cm shorter).
d. Very serious risk of bias (-2)/very large effect (+2): all four studies used an uncontrolled before-after design, but the effect is unlikely to be explained by bias alone. The effect is very large (the point estimate and both ends of the 95% CI indicate very large effects relative to the threshold, with the direction of effect reversed from usual pubertal development), so the evidence was upgraded for a very large effect (+2).
e. Very serious indirectness (-2): breast cup size is an indirect measure of pubertal development.
f. Extremely serious risk of bias (-3): all three studies use an uncontrolled before-after design and were rated at critical risk of bias for this outcome.

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License


FOR PUBLIC CONSULTATION

g. Serious imprecision (-1): the 95% CI crosses a threshold for an important effect (SMD=0.2) such that the result is compatible with important benefit (SMD= -0.25) and little to no change (SMD=0.13).
h. This finding is consistent with hypothesis prespecified in the protocol that puberty suppressants may prevent deterioration in gender dysphoria and body image and satisfaction, rather than improve these outcomes See Figure 1 in the prespecified protocol for this systematic review, which shows the underpinning logic for how these outcomes relate to the primary and secondary goals of treatment and the predicted effects.
i. Serious imprecision (-1): the 95% CI crosses a threshold for an important effect (SMD=0.2) such that the result is compatible with little to no change (SMD= -0.03) and slight, but important worsening (SMD=0.33).
j. The threshold for this outcome is set low because while this is a rare event, left undetected it can have serious consequences.
k. Serious risk of bias (-1): the single study is at moderate risk of bias due to confounding
l. Serious imprecision (-1): the 95% CI crosses a threshold for an important effect (20 events per 100000 person years), so the result is compatible with important benefit (37 fewer cases of incident IIH) and little to no difference (6 more cases). Although the result is imprecise, there is no indication of important harm (i.e. no indication of an increase in the incidence of IIH above the threshold).
m. Extremely serious risk of bias (-3): single study that uses an uncontrolled before-after design.
n. Serious imprecision (-1): Using a threshold for change of 30ms, the 95% confidence interval slightly crosses the threshold such that the result is compatible with little or no difference and important harm. This is most likely due to the small sample size.
o. Extremely serious risk of bias (-3): single cohort study rated at critical risk of bias due to missing outcome data (follow-up data available for 75/204, intervention status ascertained at follow-up, so missingness may be due to intervention and outcome).
p. Extremely serious imprecision (-3): the 95% CI crosses the threshold for an important effect (SMD=0.2), but the result is compatible with very large benefit (SMD=1.5) and little to no difference (SMD=0.03), which is extremely imprecise.
q. Very serious risk of bias (-2): two studies at serious risk of bias, and one at moderate risk of bias. One study at critical risk of bias due to missing outcome data (showing a large effect) but this study has negligible influence on the analysis (weight 4%).
r. No serious imprecision: although the CI crosses the threshold for a small benefit, the effects favour puberty suppression and the upper bound of the CI (SMD=-0.15) is close to the threshold (SMD=-0.2).
s. Very serious risk of bias (-2): Most of the included studies are at serious risk of bias, including the study with 43% of the weight in the analysis.
t. No serious imprecision: although the CI crosses the threshold for a small benefit, the effect favours puberty suppression and the upper bound of the CI (SMD=-0.11) is close to the threshold (SMD=-0.2).
u. Very serious risk of bias (-2): study with 100% of weight in the analysis rated at serious risk of bias.
v. Very serious imprecision (-2): Using a 1% threshold for an important effect, the 95% CI crosses both thresholds such that the result is compatible with important benefit and important harm.

## 4.3.2 Critical outcomes

### Pubertal development

Five studies (van de Grift 2020, Carmichael 2021; Fisher 2024, Schagen 2016; Lahaije 2025) reported results for a measure of pubertal development that contributed data for this outcome. There is evidence from these studies that puberty suppressants likely suppress pubertal development importantly when measured by penile length, menstruation and breast development. Effects on penile length and breast size were measured in adulthood, which provides evidence about longer-term effects on physical characteristics beyond immediate impacts in puberty. Results and study characteristics for each of the three outcomes follow, together with a summary of other pubertal development outcomes not addressed in the summary or synthesis.

**Penile length**

One study with 103 participants reported results for penile length (a measure of pubertal development) and contributed data to this comparison (van de Grift 2020). The study compared effects in a group that received puberty suppressants (GnRHa) to a group that did not. Separate comparisons are presented for those who commenced puberty suppressants at Tanner 2/3 and Tanner 4/5.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

*Effects*

There is moderate certainty evidence that puberty suppressants likely suppress pubertal development importantly (shorter penile length) compared to no puberty suppression (MD 3.26 cm shorter, 95% CI: 3.09 cm to 3.44 cm; 1 retrospective cohort study, 103 participants; **Figure 3**). Effects were larger among those who received puberty suppressants at Tanner stage 2/3 (MD 4.90 cm shorter, 95% CI 4.63 to 5.17 cm shorter). While the evidence was downgraded for risk of bias and indirectness, the effect is large and in the opposite direction to the trajectory of pubertal development, so very unlikely to be explained by bias or other factors. For this reason, the evidence was upgraded to moderate.

*Study characteristics*

*Population*. All participants had received gender affirming care at the gender service at Amsterdam University Medical Centre and were returning as adults (18 years or over) for assessment for gender affirming surgery (GAS).

*Interventions*. On average, participants had started puberty suppressants at age 15. The exact duration of treatment was not reported, but most started GAH 3 years after starting PS at Tanner 2/3, or 2 years after starting PS at Tanner 4/5. Both groups had received gender affirming hormones.

*Outcome measurement.* Data were collected as part of routine assessment for GAS, approximately 4 to 5 years after starting on puberty suppressants.

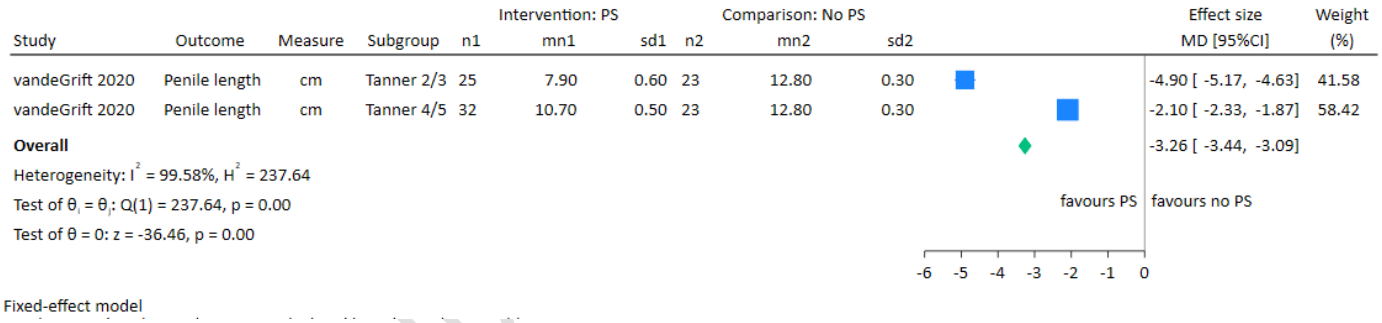



**Fig 3.** | Forest plot showing effect of puberty suppression versus no puberty suppression on pubertal development (penile length, cm). Length of follow-up: approximately 4 to 5 years from commencing puberty suppression. The comparison group n values are halved as each subgroup is compared to the same participants. MD=mean difference. PS=puberty suppression. n=number of participants. Blue lines show 95% confidence intervals (CI).

**Menstruation**

Four studies reported results for menstruation (a measure of pubertal development) and contributed data to this comparison. The studies compared effects before and after receiving puberty suppressants (GnRHa).

*Effects*

There is high certainty evidence that puberty suppressants suppress menstruation compared to no puberty suppression (proportion menstruating at follow-up was zero, 95% CI 0 to 0.01; 4 uncontrolled before-after studies, 143 participants; **Figure 4**). Before puberty suppression (baseline), 119 of 143 participants were at menarche; none of the 143 participants were menstruating at follow-up. While all evidence from uncontrolled before-after

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

studies was rated as critical risk of bias in this review, the effect on menstruation is very large and in the opposite direction to the expected trajectory of pubertal development. The effect is highly unlikely to be explained by other factors, including study limitations. For this reason, the evidence was upgraded to high certainty.

*Study characteristics*

*Population*. All participants had received gender affirming care at a gender service.

*Interventions*. On average, participants started puberty suppressants at age 14 (two studies) or Tanner stage 4 (range 2 to 5; one study). The age of commencement was not reported in one study. The exact duration of treatment was not reported. Triptorelin intramuscular injections were used in all four studies.

*Outcome measurement.* Data were collected as part of routine care at 3 to 12 months from commencing puberty suppression.

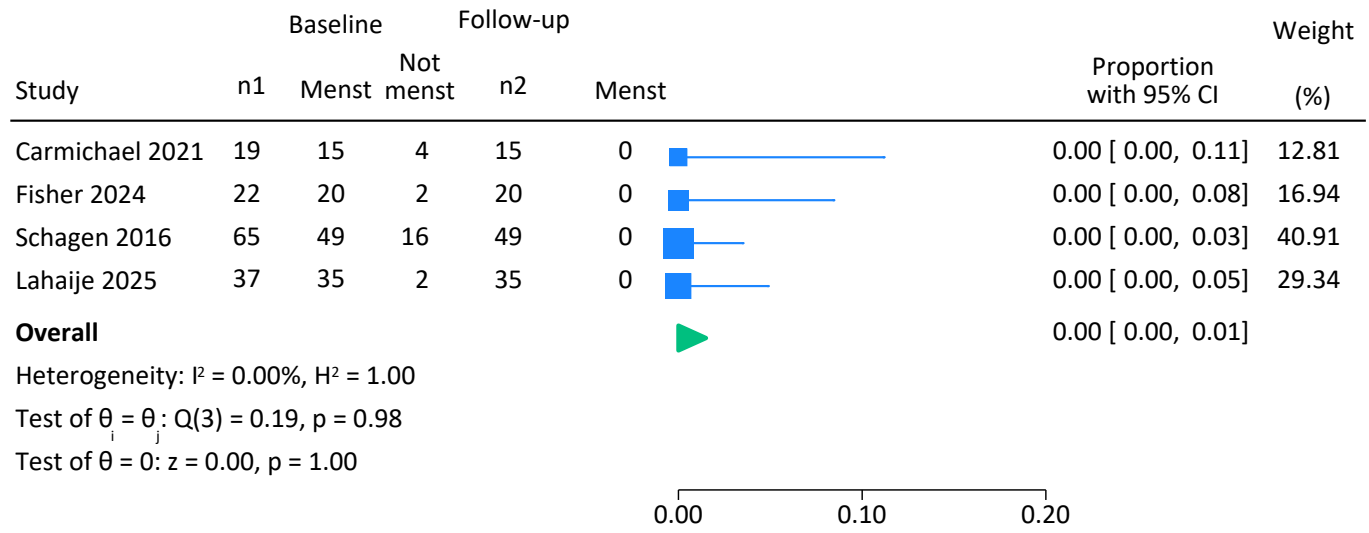



**Fig 4.** | Forest plot showing effect on pubertal development (menstruating) before (baseline) and after (follow-up) puberty suppression. Conditional proportions were calculated for those who were menstruating at baseline. Length of follow-up: approximately 3 to 12 months from commencing puberty suppression. Menst= menstruating. n=number of participants. Blue lines show 95% confidence intervals (CI).

**Breast size**

The same study that reported on penile length also reported results for breast size (measured as the proportion with a breast cup size of B or larger) and contributed data to this comparison (van de Grift 2020). The study compared effects in a group that received puberty suppressants (GnRHa) to a group that did not. Separate comparisons are presented for those who commenced puberty suppressants at Tanner 2/3 and Tanner 4/5.

*Effects*

There is low certainty evidence from this study that puberty suppressants may suppress breast development importantly (smaller breast cup size) compared to no puberty suppression (47 fewer per 100 had a breast cup size

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

of B or larger, 95% CI 32 fewer to 62 fewer; 1 retrospective cohort study, 126 participants; **Figure 5**). Effects were larger among those who received puberty suppressants at Tanner stage 2/3 (81 fewer per 100) than at Tanner stage 4/5 (see **Figure 5**). While the evidence was downgraded for study limitations (risk of bias) and the indirectness of breast cup size as a measure of pubertal development, the effect is very large and in the opposite direction to the trajectory expected in pubertal development, so very unlikely to be explained by bias or other factors. For this reason, the evidence was upgraded to low.

*Study characteristics*

Described with results for penile length.

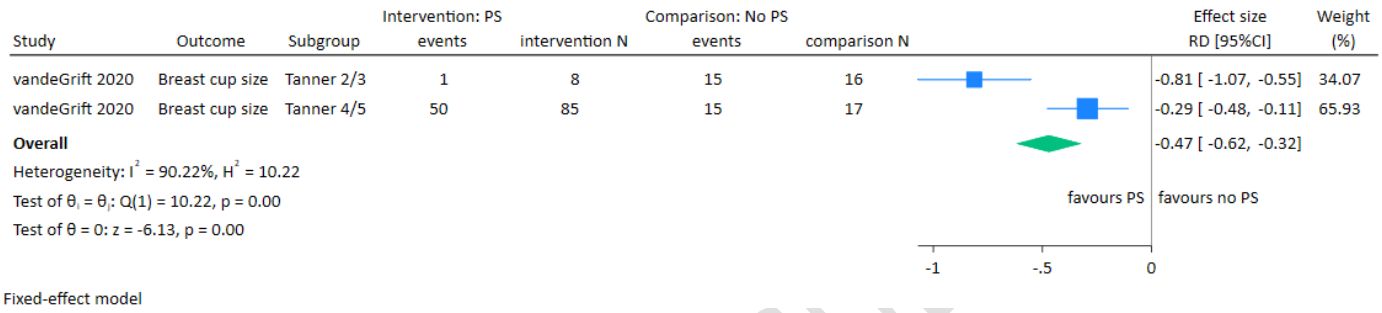



**Fig 5.** | Forest plot showing effect of puberty suppression versus no puberty suppression on pubertal development (breast size). The event is defined as developing a breast cup size of B or larger. Length of follow-up: approximately 4 to 5 years from commencing puberty suppression. The comparison group n values are halved as each subgroup is compared to the same participants. RD=risk difference. PS=puberty suppression. N=number of participants. Blue lines show 95% confidence intervals (CI).

**Other pubertal development outcomes**

Tanner stage was reported as an outcome (rather than as a baseline characteristic) in one study (Fisher 2024); however, the effect estimates could not be calculated from this data (presented in a violin plot without the required statistics). Testes volume was measured in two studies, one of which did not report results (Lahaije 2025) and the other reported incompletely (Schagen 2016; the data could not be used to calculate effect estimates). Additionally, van de Grift 2020 reported on testicular descent but these results were not selected for analysis (in line with our prespecified plan to select one result per study for analysis in each outcome domain).

## Gender dysphoria

Three studies (100 participants) reported results for gender dysphoria and contributed data to this comparison. The studies compared effects before and after receiving puberty suppressants (GnRHa).

*Effects*

Results for gender dysphoria are presented in the summary of findings table and in **Figure 6** (forest plot). Little to no change in gender dysphoria was observed between commencing PS and follow-up (SMD 0.06 lower, 95% CI 0.25 lower to 0.13 higher; lower is better; $I^2 = 0\%$, 1 uncontrolled before-after study, 100 participants; very low certainty). However, the evidence is of very low certainty due to extremely serious risk of bias (arising from the use of an uncontrolled before-after design in all three studies) and imprecision. Without a control group, it is not

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

possible to infer whether the absence of change in gender dysphoria was due to the stabilising effects of puberty suppression (as per the hypothesis in the protocol) or other factors.

*Study characteristics*

*Population*. All participants had received gender affirming care at a gender service.

*Interventions*. On average, participants started puberty suppressants at age 14 and were treated for 2 years. Triptorelin, intramuscular injections were used in one study and the GnRHa was not reported in the other two.

*Outcome measurement.* Data were collected as part of routine care at 12 months from commencing puberty suppression in two studies, and 12 months to 2 years in the third.

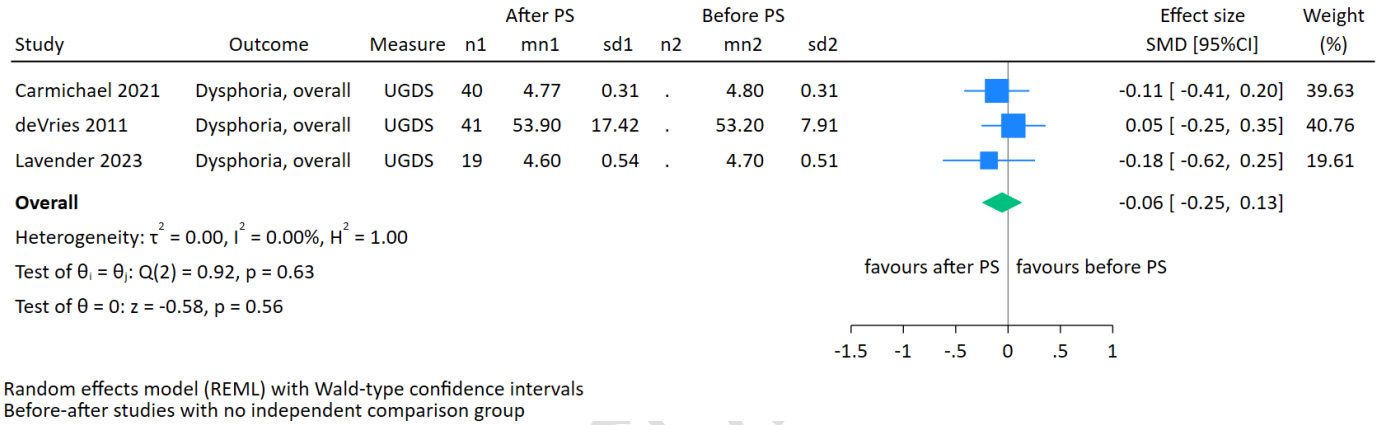



**Fig 6**. | Forest plot showing effect on gender dysphoria before (baseline) and after (follow-up) puberty suppression Length of follow-up: approximately 12 months from commencing puberty suppression. PS=puberty suppression. SMD=standardised mean difference. n=number of participants. mn=mean. sd=standard deviation. UGDS=Utrecht Gender Dysphoria Scale. Blue lines show 95% confidence intervals (CI). [7]

## Body image and satisfaction

The same three studies that reported on gender dysphoria also reported results for body image and satisfaction and contributed data to this comparison. The studies compared effects before and after receiving puberty suppressants (GnRHa). Study characteristics are the same as for gender dysphoria.

*Effects*

Results for body image and satisfaction are presented in the summary of findings table and in **Figure 7** (forest plot). Little to no change in body image and satisfaction was observed between commencing PS and follow-up (SMD 0.15 higher 95% CI 0.03 lower to 0.33 higher; lower is better; $I^2$ = 0%, 1 uncontrolled before-after study, 118 participants; very low certainty). However, the evidence is of very low certainty due to extremely serious risk of

[7] We cannot rule out the possibility that data from at least some of the same participants may have been included in both Carmichael and Lavender. Participants in Carmichael were recruited between April 2011 and April 2014 and commenced GnRHa treatment between June 2011 and April 2015. The timing of treatment is unclear in Lavender, with participants reported as 'attending' the clinic between 2014 and 2018. We undertook a sensitivity analysis removing Lavender from the MA which did not importantly change the effect estimate.

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

bias (arising from the use of an uncontrolled before-after design in all three studies) and imprecision. Without a control group, it is not possible to infer whether the absence of change in body image and satisfaction was due to the stabilising effects of puberty suppression (as per the hypothesis in the protocol) or other factors.

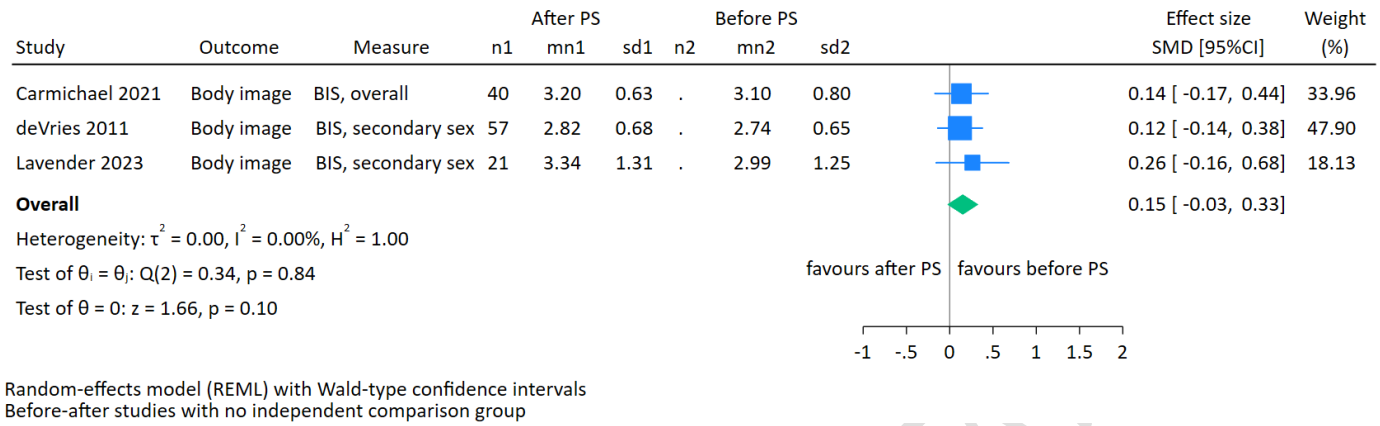



**Fig 7.** | Forest plot showing effect on body image and satisfaction before (baseline) and after (follow-up) puberty suppression. Length of follow-up: approximately 12 months from commencing puberty suppression. PS=puberty suppression. SMD=standardised mean difference. n=number of participants. mn=mean. sd=standard deviation. BIS=Body Image Scale. Blue lines show 95% confidence intervals (CI).[1]

### Serious adverse events

Two studies reported on adverse events of special interest – idiopathic intracranial hypertension (IIH) and QTc prolongation. A third study reported outcomes for people who undergo future gender affirming surgeries. We also screened studies for spontaneously reported adverse events and did not find any serious adverse events reported.

#### Idiopathic intracranial hypertension (IIH)

The study on idiopathic intracranial hypertension (a serious but rare adverse outcome), used population-wide data from several national databases in Sweden. The study compared the incidence of IIH among people with a recorded diagnosis of gender dysphoria who had received GnRHa to those who had not (Karamanis 2023). There was low certainty evidence (due to risk of bias and imprecision) that GnRHa may have little to no effect on the incidence of idiopathic intracranial hypertension compared to no GnRHa (Rate ratio: 15.7 fewer IIH events per 100000 person-years, 95% CI 37.4 fewer to 6.1 more; 1 non-randomised study, 4230 people, low certainty evidence) (see **Table 4**). No cases of IIH were identified among individuals receiving GnRHa (410 people, median follow-up 1.4 years).

#### QT interval prolongation

The study on QTc interval was an uncontrolled before-after study involving 16 people (Zaitoon 2025). A prolonged QTc interval is a risk factor for serious heart rhythm abnormalities. There was very low certainty evidence (due to extremely serious risk of bias and imprecision) that puberty suppressants (GnRHa) may have little to no effect on QTc interval. Results were interpreted using a threshold commonly used by drug regulators (see end of **Table 4**).

FOR PUBLIC CONSULTA

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

### Outcomes from future gender affirming surgeries

One study compared post-surgical outcomes among those who received PS to those who did not (van de Grift 2020). The outcome descriptions were presented to the Committee to seek their views on whether any of the outcomes should be considered under the serious adverse events domain. No information other than the outcome measures (e.g. breast tissue volume resected) and the comparison was provided (i.e. not study identifiers, results, or other data). The Committee agreed that evidence about volume of breast tissue resection should be considered under SAEs.

There was moderate certainty evidence (due to risk of bias) that, for those undergoing future gender affirming top surgery, puberty suppressants probably result in a large reduction in the amount of breast tissue removed or resected (**Figure 8**). Effects were largest among those who received puberty suppressants at Tanner stage 2/3 (MD 540 grams less tissue resected, 95% CI 605 lower to 475 lower).

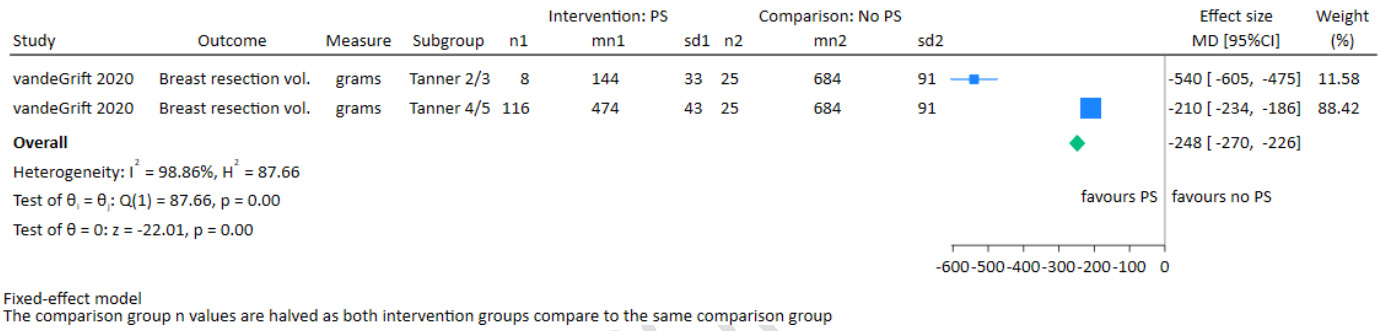



**Fig 8.** | Forest plot showing effect of puberty suppression versus no puberty suppression on breast tissue resection volume (grams). Length of follow-up: approximately 4 to 5 years from commencing puberty suppression. MD=mean difference. PS=puberty suppression. n=number of participants. mn=mean. sd=standard deviation. Blue lines show 95% confidence intervals (CI).

### Other reports of adverse events

In screening the full text of studies, we identified 10 studies (11 papers) that considered adverse events, meaning that the study either explicitly stated that 'adverse events' were an outcome of interest, or we judged that data relevant to adverse events was reported.

Five of the 10 studies were included in the review:

- 3 are reported under serious adverse events (Karamanis 2023, Lahaije 2025, van de Grift 2020).
- 2 reported on adverse events, including those that were spontaneously reported (Carmichael 2021 and Lahaije 2025), none of which met the commonly used definition for SAEs that we adopted for the review (**Appendix 1a**, outcomes table).

Of the five studies that were ineligible for the review:

- 2 reported cross-sectional data (as is usual when an outcome that cannot be measured prior to intervention is reported from an uncontrolled before-after study). In one (Olson 2024), none of the

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

reported AEs met our definition of serious and the other reported two cases of non-serious AEs that resolved (Khatchadourian 2014).

- 3 did not report separate data for received puberty suppression, so effects could not be isolated from other interventions (GAH, antiandrogens, 'any intervention').

Information about non-serious adverse events was summarised as additional information in the GRADE evidence to decision framework for the Committee to consider.

### 4.3.4 Important outcomes

For important outcomes, studies had to provide data from a comparison of outcomes among a group who had received PS compared to a group who did not receive PS. Comparisons of before and after treatment with puberty suppression were not eligible.

#### Satisfaction with care

We did not identify any study that met the eligibility criteria for this outcome. Several studies reported cross-sectional data on satisfaction with care involving puberty suppression (after treatment, without comparison to a group that did not receive puberty suppression). It is not possible to isolate the effects of puberty suppression from other components of care. Instead, these data were considered as additional information in the GRADE evidence to decision framework.

#### Quality of life

One study (32 participants) reported data on quality of life comparing effects in a group that received puberty suppressants (GnRHa) to a group that did not (Becker-Hebly 2021).

**Effects**

The evidence on quality of life is of very low certainty due to risk of bias and imprecision, that make the results too uncertain to interpret (**Figure 9**). The study is at extremely serious risk of bias because of missing outcome data (data for 75/204 participants were available), and extremely serious imprecision because the 95% confidence interval is compatible with both 'little to no effect' and a 'very large' effect.

**Study characteristics**

*Population*. All participants had received gender affirming care at a gender service.

*Interventions*. The age at which participants started puberty suppression was 15 years on average. The exact duration of treatment was not reported, nor was the type of GnRHa.

*Outcome measurement.* Data were collected as part of routine care 6 months to 4 years from commencing puberty suppression.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

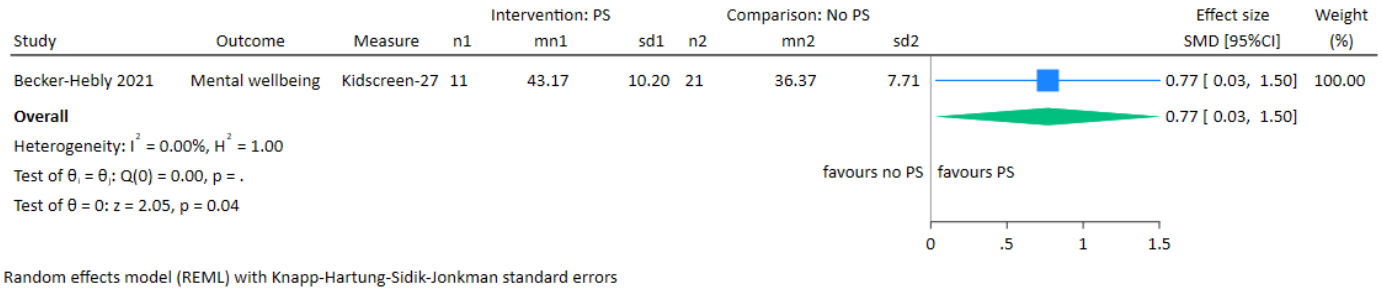



**Fig 9.** | Forest plot showing effect of puberty suppression versus no puberty suppression on quality of life and wellbeing. Length of follow-up: 6 months to 4 years from commencing puberty suppression. PS=puberty suppression. SMD=standardised mean difference. n=number of participants. mn=mean. sd=standard deviation. Blue lines show 95% confidence intervals (CI).

## Social functioning

Four studies (991 participants) compared effects on social function in a group that received puberty suppressants (GnRHa) to a group that did not.

### Effects

There is low certainty evidence (due to risk of bias) that puberty suppressants may improve social functioning (overall psychosocial, social) compared to no puberty suppressants (SMD 0.30 lower, 95% CI 0.45 lower to 0.15 lower; lower is better; $I^2$ = 0%, 4 cohort studies, 991 participants; low certainty; **Figure 10**; **Table 4**). All four studies used a cohort design with an independent comparator group.

### Study characteristics

*Population*. All participants had received gender affirming care at a gender service.

*Interventions*. The age at which participants were reported to have started puberty suppressants was not reported in one study and varied in the other 3 studies (on average, 12 years, 15 years, and 16 years). The duration of treatment was 18 months in one study and was not reported in the other 3. None of the studies reported the type of GnRHa, route of administration or treatment schedule.

*Outcome measurement.* Data were collected as part of routine care at 6 months to 4 years from commencing puberty suppression in one study, 18 months in 2 studies, and was not reported in the fourth study.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

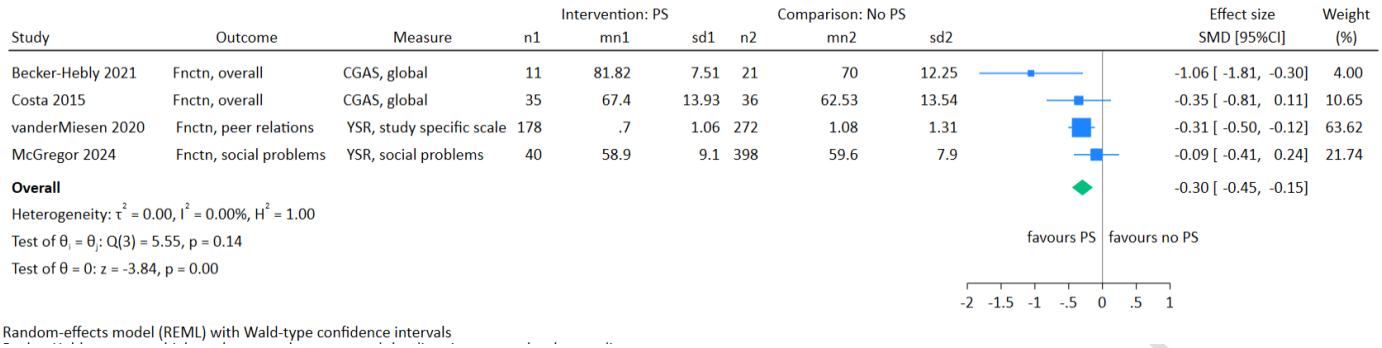



**Fig 10**. | Forest plot showing effect of puberty suppression versus no puberty suppression on social functioning. Length of follow-up: approximately 18 months from commencing puberty suppression (length of follow-up in studies with the most weight in the analysis). MD=mean difference. PS=puberty suppression. n=number of participants. mn=mean, sd=standard deviation. Blue lines show 95% confidence intervals (CI). CGAS=Children's Global Assessment Scale, YSR=youth self-report. Scores were reversed for CGAS, so that all scales operate in a consistent direction (lower scores are better).

## Mental health – psychological functioning

Four studies (4414 participants) compared effects on mental health – psychological functioning in a group that received puberty suppressants (GnRHa) to a group that did not.

### Effects

There is low certainty evidence (due to serious risk of bias) that puberty suppressants may improve psychological functioning compared to no puberty suppressants (SMD 0.39 lower, 95% CI 0.68 lower to 0.11 lower; lower is better; $I^2$ = 30.70%, 4 cohort studies, 4414 participants; **Figure 11**; **Table 4**). All four studies used cohort design with an independent comparator group.

### Study characteristics

*Population*. In three studies, participants had received gender affirming care at a gender service. The fourth study was a large national survey in the US where other care received was likely to vary widely because of the diversity of contexts and circumstances in which respondents lived.

*Interventions*. The average age at which participants were reported to have started puberty suppressants was not reported in one study, 12 years in a second, 15 years in a third study, and around 15 years in the national survey. The duration of treatment was 18 months in one study and was not reported in the other 3. None of the studies reported the type of GnRHa, route of administration or treatment schedule.

*Outcome measurement.* Data were collected as part of routine care at 6 months to 4 years from commencing puberty suppression in one study, 18 months in one study, and was not reported in a third study. In the national survey, participants had started on puberty suppressant about 6 years prior to completing the mental health measure (Kessler Psychological Distress Scale).

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

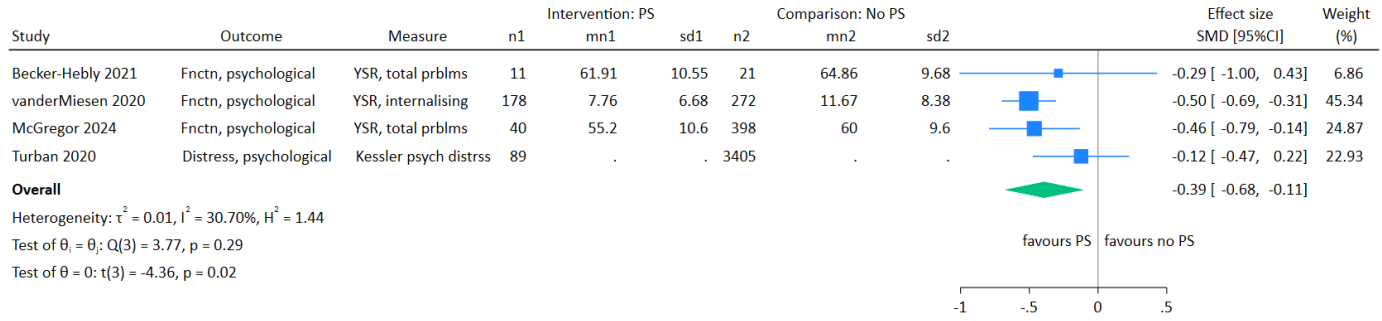



**Fig 11**. | Forest plot showing effect of puberty suppression versus no puberty suppression on mental health – psychological functioning. Length of follow-up: approximately 6 months to 6 years from commencing puberty suppression. MD=mean difference. PS=puberty suppression. n=number of participants. mn=mean. sd=standard deviation. Blue lines show 95% confidence intervals (CI). Kessler= Kessler Psychological Distress Scale, YSR=youth self-report.

### Mental health – suicidality diagnosis

One study (3321 participants) compared effects on mental health – suicidality among young people who had received puberty suppressants (GnRHa) to those who had not (Nunes-Moreno 2025). The study used health record data from PEDSnet, a research network of 5 paediatric hospitals in the United States, to identify trans and gender diverse young people with a diagnosis of suicidality made during emergency department (ED) or inpatient visits.

#### Effects

There is very low certainty evidence that puberty suppressants may reduce the incidence of suicidality diagnoses in the ED (HR 0.79, 95% CI 0.47 to 1.31; 1 study 3321 participants). The evidence is very uncertain due to very serious risk of bias and very serious imprecision (**Table 4**). In absolute terms, this translates to 3 fewer diagnoses per 100 people over 12 months (from 8 fewer to 4 more; using a baseline risk from the Trans Pathways study – see **Table 4** footer for details).

#### Study characteristics

*Population*. Young people who had an ED or inpatient visit, and at least one outpatient visit at one of the 5 hospitals with data in PEDSnet.

*Interventions*. Puberty suppressants were started at a mean age of 12.8 years (mean; SD 2.1). The duration of treatment was not available, nor was the type of GnRHa, the route of administration or treatment schedule.

*Outcome measurement.* Outcome data came from medical records reporting suicidality diagnoses at an ED or inpatient visit. On average, those included in the study sample had been on puberty suppressants for 2 years.

### Mental health – death by suicide

We did not find any eligible studies that examined death by suicide.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

### Adverse events of special interest – bone health

We did not find any eligible studies that examined bone health – fractures or functional outcomes. Additional information from a recent systematic review and findings from screening studies for this review is reported in the Discussion.

### Adverse events of special interest – fertility

We did not find any eligible studies that examined fertility potential (addressed in the Discussion). .

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

# 5 Discussion

## 5.1 Summary of main results

We included 15 non-randomised studies that examined the effects of puberty suppression on outcomes for young people experiencing gender incongruence or dysphoria. The studies reported data from 12,756 people.

We found evidence that compared to usual care, GnRHa is effective for *suppressing puberty*, with large effects on menstruation (high certainty), penile length (moderate certainty), and breast size (low certainty). The usual trajectory of pubertal development is predictable, and hence observing large effects in the opposite direction increased our certainty that the effects were due to the use of GnRHa and not other factors (i.e. concerns about bias were ameliorated). Some concerns about the indirectness of penile length and breast size as measures of pubertal development explain the lower rating of the certainty of evidence for these outcomes. For those who receive subsequent GAH, GnRHa may prevent permanent physical changes with implications for future gender affirming surgeries. Specifically, large reductions in the amount of breast tissue resected were observed among those who had received GnRHa compared with those who had not (moderate certainty). This outcome was included as a measure of serious adverse events, reflecting the seriousness of invasive mastectomy.

We found low certainty evidence that there may be small but important improvements in *mental health (psychological functioning)* and *social functioning (reduction in overall problems and social problems)* for those who receive GnRHa compared to those who do not (low certainty evidence). From another study that reported data that could not be combined with other results for mental health, we found evidence that GnRHa may lead to small but important reductions in *suicidality diagnoses* in inpatient and emergency department settings, but the evidence is very uncertain. While the study was well designed (reflected in low risk of bias on most domains), the results were imprecise and there were differences in characteristics between the groups compared (those who received GnRHa and those who did not), which may impact the outcome and could not be completely controlled for.

The effects on *gender dysphoria* and *body image satisfaction* were very uncertain because the evidence came exclusively from studies that compared these outcomes before and after receiving GnRHa. While little to no change in gender dysphoria and body image was observed between commencing puberty suppressants and follow-up, the absence of a control group means it is not possible to infer whether this was due to the stabilising effects of puberty suppression (as per the hypothesis in the logic model for the review, **Figure 1**) or other factors.

We did not find any eligible studies that reported on *death by suicide* or *satisfaction with care* (in relation to receipt of puberty suppression). One small study examined effects on *quality of life and wellbeing*, but the evidence was too uncertain to interpret (due to the observed effect being compatible with little to no effect and large benefit, i.e. imprecise).

We did not find evidence of *serious adverse events* with GnRHa, noting that follow-up was short term. We identified two studies that examined serious adverse events that are commonly evaluated for drugs to identify safety concerns or signals that active monitoring may be needed in routine use. A large population-level study provided evidence that GnRHa may have little to no effect on the incidence of idiopathic intracranial hypertension (low certainty), which is a serious rare adverse outcome. One small study (16 participants) found that GnRHa may have little to no effect on QTc interval, but this evidence is very uncertain. We did not find any eligible studies that examined *bone*

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

*health – fractures or functional outcomes*, a result which we discuss in the context of a recent systematic review (see 'Agreements and disagreements with other systematic reviews'). Nor did we find any eligible studies that examined *fertility potential*.

## 5.2 Limitations of the evidence included in the review

Limitations of the evidence in a review can relate to the alignment (relevance) of the questions addressed in primary studies and factors that influence the certainty of evidence (e.g. design limitations of primary studies). We address both points, noting that the two are interrelated in this review. Reporting of studies is considered throughout because it has cross-cutting implications.

Relevance relates to how closely the study population(s), intervention(s), comparator(s) and outcomes (i.e. PICO) match the questions specified for the review. Overall, there was good coverage of the review questions and what the included studies addressed in terms of relevant populations, interventions and outcomes. There were some limitations, that if addressed, could enhance the evidence base for decision-making. Very few studies reported results separately by Tanner stage (or age) except those that reported on pubertal development. There was also limited reporting by gender (or sex assigned at birth). Reporting this data may help in understanding whether puberty suppression is more effective for some groups than others. Reporting of equity-related population characteristics was limited, so it is unclear to what extent studies included groups that may experience greater inequity due to intersecting factors such as place of residence (especially rural and remote), neurodivergence, First Nations identity, ethnic or racial background or disability. More recent studies tended to provide more complete reporting of equity-related characteristics. None of the included studies were in Australia, so there is no representation of Aboriginal or Torres Strait Islander peoples.

Several prioritised outcomes were not addressed in included studies. From our screening of 145 full text articles (Supplement 2), we found 27 studies that reported on bone health (bone mineral density [BMD] in almost all) but only two of these reported on fractures (neither used an eligible design; both reported that no fractures occurred) (101, 102). Given the relevance of bone health, it is likely that data on fractures is collected as part of periodic health checks (from which data are sought for studies), so non-reporting of this outcome is a surprising omission. We found only cross-sectional data on fertility and satisfaction with care. Only one eligible study measured quality of life. Some outcomes are more commonly reported - especially mental health outcomes - but often in studies with an ineligible design. This points to limitations in the evidence arising from study design.

The choice of comparators is, perhaps, the most important study design consideration in this review. Limitations arise from comparison of outcomes before and after GnRHa without an independent comparator group. While we restricted inclusion of data from uncontrolled before-after studies to critical outcomes (including pubertal development which has a clear trajectory, strengthening inferences from the design), the evidence on gender dysphoria and body image was of very low certainty because it came entirely from uncontrolled before-after studies. Results from these studies were judged to be at critical risk of bias due to confounding. Without a comparator group, it is generally not possible to know whether any change in the outcome is due to GnRHa, other interventions an individual receives (such as psychosocial support), or other factors (including temporal changes). In other systematic reviews, this has ramifications for mental health outcomes because, for example, adolescence is associated with the onset of common mental disorders (depression, anxiety) resulting from multiple biological

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

and social factors (103, 104). Attributing any change in mental health outcomes during adolescence to GnRHa would be problematic in the absence of a comparator group that does not receive GnRHa. While acknowledging that the choice of comparator requires careful consideration, there may be an opportunity to strengthen the design of studies using existing data. We noted that suitable data may have been available to some investigators because they used medical records to identify outcomes for those who had received only puberty suppression and excluded those who had not. In other studies, data were available but were not reported separately for those who had received puberty suppression (i.e. the study reported results for a combined group that included youth who had either received only puberty suppressants, only GAH or both). To some extent, these decisions reflect a difference in the question addressed by the study authors and our review, but the value of studies can be maximised by reporting data in formats that enable reuse in evidence synthesis.

Study design limitations were the main reason for lowering the certainty of evidence, but it is important to recognise that not all study design limitations are preventable, and others are difficult to control. As is near universal for reviews that include non-randomised studies, risk of bias due to confounding was common and reduced the certainty of evidence. This does not lessen the importance of non-randomised studies for addressing the questions in this review, nor does it imply that the studies were not well-designed. Where multiple studies were available, the findings were consistent. Most studies provided direct evidence to address the question. For outcomes for which there were fewer participants contributing data or for rare events (some of the serious adverse events), there was imprecision, but generally this reduced certainty in the size of effect not the direction (with most effects favouring puberty suppression, except for body image and satisfaction).

Finally, many commonly used scales for measuring outcomes related to gender identity, expression and dysphoria are older measures that reflect outdated beliefs and terminology (including pathologising trans and gender diverse identities) (16, 39). While there is evidence of the psychometric properties of these measures, recent evaluations raise important concerns that suggest a lack of face and content validity for measuring outcomes of importance such as gender dysphoria and body image.

## 5.3 Limitations of the review processes

We followed Cochrane and GRADE guidance to minimise preventable bias in the review process and specified our methods ahead of commencing screening, however limitations may exist.

Rather than conduct our own searches (except for running an update search), we sourced studies that had been included in recent systematic reviews and evidence reports of adolescents with gender incongruence or dysphoria. This approach avoided rerunning searches and rescreening records that had already been screened many times over by different review teams. Although this approach relies on the comprehensiveness of the search methods of other reviews, given the large number of reviews included and the wide array of sources searched, we benefited from the collective expertise of information specialists who had developed search strategies across different databases and sources, thus maximising the pool of available studies and minimising the risk of missing relevant ones.

Determining eligibility of non-randomised studies requires careful inspection of design features (40, 41) to avoid missing potentially eligible studies or misinterpreting the study design. To address this, we retrieved a higher

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

proportion of full-text studies to ascertain the study design than would be typical for reviews that rely on the investigators labelling of designs. We screened the full text of all studies listed as excluded by previous reviews (Supplement 2). We also implemented a triage step in our data extraction where we coded design features of near miss studies, including those where the description of the design or the availability of data on puberty suppression required close inspection.

While the requirement to consider puberty suppression and GAH in sequential systematic reviews enabled a more streamlined approach to the conduct and reporting of the puberty suppression review, the evidence about these two treatments is closely interlinked. This interlinking has both conceptual and practical implications. For example, a complete understanding of the impact of puberty suppression on bone health requires evidence up to when peak bone mass is reached. The most relevant evidence spans from adolescence into early adulthood, which means that those who have received puberty suppression will have had subsequent exposures from interventions and lifestyle factors that are associated with bone health and peak bone mass. GAH is one of these exposures. To gain a complete understanding of effects, puberty suppression needs to be considered in the context of GAH (and other exposures subsequent to puberty suppression). Such questions could be integrated into the review of GAH. On a practical level, the close interlinking of evidence on puberty suppression and GAH complicates identification of relevant evidence. Although uncommon, we found studies of GAH with no mention of puberty suppression in the title or abstract that reported data suitable for estimating effects of puberty suppression independently of GAH. Locating this data required scrutiny of results tables and text, so it is possible that we may have missed eligible data. However, we screened the full text of studies excluded from other reviews of puberty suppression, several of which had considered GAH in tandem.

We did not contact authors to obtain missing data or confirm assumptions made in synthesizing the data (e.g. the correlation between baseline and follow-up measures). The assumptions only applied to results for gender dysphoria and body image, the interpretation of which did not change based on sensitivity analyses and the findings are of very low certainty.

Each stage of the review has been informed by the expertise of people with lived experience, members of the Committee and members of the Reference Panel. However, the review team does not include authors who identify as trans, have previous experience undertaking research on trans health, or are health professionals caring for people who are trans and gender diverse. We recognise that this influences our worldview, and interpretation of the evidence.

## 5.4 Agreements and disagreements with other studies or reviews

Multiple reviews have examined the effects of puberty suppression for young people experiencing gender incongruence. We examined the methods and findings of 11 published reviews of the effectiveness and safety of puberty suppression. These reviews do not reflect the full extent of reviews on puberty suppression, but they include prominent reviews (including the review that informed the Cass review in the UK (32), a review by Miroshnychenko for the Society for Evidence-based Gender Medicine (SEGM) (29), a review by Rand Health Care (26), a review by an Italian team with funding from the Italian Ministry of Health (34), a review by the Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU) (28) and a review for the Utah Department of Health (27)). We also considered reviews conducted for jurisdictions in Australia and New Zealand

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License


FOR PUBLIC CONSULTATION

(i.e. for New South Wales Health (25), Queensland Health (30), and the New Zealand Ministry of Health (23, 24)). In addition to the 11 reviews, we consider a systematic review published in 2026 on the effects of GnRHa with or without subsequent GAHT on bone accrual (105).

Of the 11 reviews we examined, we did not find any that described a comprehensive process to prioritise outcomes, nor did we find mention of whether the outcomes considered reflected the priorities of those with lived experience. For commissioned reviews, it is possible that this process was undertaken and reported separately. We did not find an explicit rationale for the included outcomes or consideration of anticipated effects. Many of the reviews included a large number of outcomes, some of which had been rated as of limited importance in the prioritisation process for this review. Both Cochrane and GRADE recommends against including a large number of outcomes (especially those of limited importance) because this "risks overwhelming and potentially misleading readers" (46). The reviews vary widely in terms of the questions addressed (and associated PICO eligibility criteria) and their methodology.

Most notable differences are in the:

- eligibility of study designs without an independent comparator group (uncontrolled before-after designs were included for all outcomes, without rationale for including the design for outcomes such as mental health ; some also included cross-sectional data, qualitative data or both),
- eligibility of studies in which some or all recipients of puberty suppression had also received GAH prior to follow-up of outcomes (precluding isolation of effects of GnRHa if separate outcome data are unavailable),
- use of meta-analysis (used in one review; results from existing meta-analyses may have been interpreted in 2 overviews),
- methods used to interpret study-level results and synthesise results when meta-analysis is not used (fully reported and appropriately used in one review; unclear in most others),
- inclusion of results from existing systematic reviews (done in two reviews),
- use of GRADE methods (applied in accordance with GRADE guidance in 4 reviews, misapplied in 2).

Below, we summarise agreements and disagreements with the two systematic reviews that report the most comparable methodologies (26, 29) and the systematic review that informed the Cass review (32). We then consider the recent review of bone health (105).

The review by Miroshnychenko 2025 uses meta-analysis and GRADE methods, suggesting it is similar methodologically. There are four studies in common with our review; the difference in the number of studies in common being partly due to the outcomes considered (of their 10 included studies, 5 measured bone health). Outcomes that were not reported by Miroshnychenko 2025 included pubertal development, body satisfaction, serious adverse effects, and quality of life. The analyses of gender dysphoria and social functioning (described as 'global functioning') are the most comparable to our analyses in terms of the outcomes considered, although the authors report different effect measures and have fewer studies in each analysis. The results for gender dysphoria (which included 2 of our 3 studies) and social functioning (which included 2 of our 4 studies) appear similar in terms of direction of effect. However, the authors interpretation of findings is not directly comparable. Results are interpreted in relation to the null rather than a threshold for a minimally important effect, such that Miroshnychenko and colleagues describe potentially trivial results as an increase or reduction in an outcome (for

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License




FOR PUBLIC CONSULTATION

example, an SMD of 0.01 is described as ‘gender dysphoria may be lower’). The authors elected not to provide a plain language interpretation for very low certainty results in their GRADE table (such interpretations are essential if results are to be used for decision-making (106) and are consistent with GRADE guidance (107)). In both approaches used by the authors (interpretation or lack thereof), the approach could lead to misinterpretation of the findings wherein effects that are large enough to be important could be ignored and trivial effects could be overemphasised.

The review by Dopp 2025 for Rand Health Care, addressed effects of puberty suppression as one of multiple questions. There are no meta-analyses, but the synthesis methods and approach to interpretation are completely reported, with a clear rationale for decisions. There were few restrictions on study designs eligible for the review, but the synthesis was limited to designs that were “sufficiently low risk of bias to justify summarising a single study” (Joanna Briggs Institute risk of bias tools were used). The authors recorded all relevant outcomes, but only summarised results for outcomes reported in multiple studies. Of the outcomes eligible for our review, Rand and colleagues report summary results with GRADE ratings for pubertal development (a ‘decrease in pubertal changes’, 20 studies, low certainty), gender dysphoria (‘improved’, 11 studies, very low certainty), mental health symptoms (‘improved mental health functioning’, 21 studies, very low certainty), fertility (‘no effects on fertility’, 2 ‘case studies only’) and side effects (‘low side effects and complications’, 18 studies, very low certainty). The number of studies contributing to each outcome reflects a lenient approach to inclusion of study designs that are at critical risk of bias (i.e. uncontrolled before-after designs) which likely explains the lower certainty of evidence for some outcomes. It also illustrates why Cochrane and GRADE recommend restricting to designs that are more suited to making causal inferences about intervention effects, and for which certainty in the evidence will therefore be higher (40, 84, 108). Conceptually, Cochrane and GRADE’s approach is analogous to limiting to randomised trials where these are feasible, ethical and suitable for addressing the outcomes of interest (108).

The review by Taylor 2024 is similar to the Rand review in its inclusive approach, but not the methodology. The synthesis approach is not reported, interpretation appears likely to be based on statistical significance, which becomes problematic especially when results from multiple studies are synthesised (see (109)), and study ‘quality’ is reported based on risk of bias assessments (using the Newcastle Ottawa Scale) without GRADE assessments. For these reasons, we have not compared the results to those in our review.

A systematic review published in 2026 by Tienforti and colleagues examined the effects of GnRHa with or without subsequent GAH on bone accrual (BMD, aBMAD and z-scores) in TGD adolescents (105). The review included 10 studies, all of which were identified in our original search (from a total of 27 excluded studies that reported on bone parameters). All 10 studies were excluded from our SR (most because of study design, not reporting an eligible outcome or both). Two of the 10 studies reported on fracture (101, 102), one of which used 6 monthly lateral spine radiographs to monitor for vertebral fractures(102), which are diagnostic for paediatric osteoporosis (110). Both studies were ineligible for our SR because they used an uncontrolled before-after design, but eligible for the review by Tienforti and colleagues. Despite this, the SR by Tienforti 2026 reported that “no cohort reported fracture outcomes” (105), p3. The results may have been overlooked because the studies did not fully report quantitative data.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

# 6 Authors' conclusions

## 6.1 Implications for practice

Although it is usual to report on implications for practice in a systematic review, we have not done so here since this falls within the remit of the Gender Guidelines Development Committee.

## 6.2 Implications for research

The design of future studies on puberty suppression has been debated, often using arguments that neglect design features that can enable inferences about causal effects from observational designs and that overstate the potential of randomised trials to resolve uncertainty (111). The ethical and equity implications of withholding puberty suppression in order to randomise to treatment are such, that non-randomised studies (observational evidence) will likely remain the most important evidence source for understanding the effects of puberty suppression. Non-randomised studies are an essential design for evaluating interventions when a trial is unfeasible or unethical (108), and enable outcomes to be followed over a longer time, which is important for serious and rare adverse events. Indeed, the historic focus in systematic reviews on randomised trials at the exclusion of non-randomised studies has been particularly problematic for understanding harms (108, 112).

Future studies need to adopt non-randomised designs and methods of analysis that provide the best possible evidence on the effects of interventions (i.e. addressing preventable sources of bias). Among newer designs, the use of target trial emulation studies offers opportunity to use routinely collected data to estimate causal intervention effects (44, 113). In a target trial, observational analyses are designed to "explicitly emulate a hypothetical randomized trial that would answer the question at hand" (44). There is precedent for use of this design in research evaluating effects of GAH for young trans people (114). While there is an argument for prospective designs, in which participants are recruited into the target trial emulation study at the start of treatment and their outcomes followed over time, such studies can be conducted using data from existing medical records. These kinds of design decisions need to carefully weigh the advantages and disadvantages of using substantial existing data to provide evidence in the shorter term ('retrospective' data), with the those of prospective designs. At a minimum, the inclusion of appropriate comparator groups and controlling for confounding needs to be prioritised to address preventable sources of bias.

There is a need for research to establish which outcomes matter most to young trans and gender diverse people, especially those considering puberty suppression and future gender affirming medical care. Systematic reviews and guidelines must prioritise outcomes that matter most to the people affected by the treatment being evaluated (i.e. 'patient' or 'people' important outcomes). To prioritise outcomes, we relied heavily on experts with lived experience, the Committee, and the Reference Panel for advice on the primary and secondary treatment goals of puberty suppression and the value placed on different outcomes. This informed the logic model for the review (Figure 1) and the outcome priorities, which differ from those addressed in numerous other systematic reviews. Empirical evidence would increase confidence in the relative importance of different outcomes, ensuring outcomes are weighed appropriately when making judgements about the balance of benefits and harms. The

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

development of core outcome sets – reflecting consensus views – could complement this research and provide guidance for primary studies on outcome measurement.

A related consideration is the emerging evidence about the appropriateness of commonly used scales for measuring gender-related outcomes (15, 16, 39). Recent studies have evaluated instruments commonly used in studies involving trans and gender diverse young people to measure patient reported outcome measures (PROMS). A central finding from these reviews is that existing measures do not adequately represent (and sometime misrepresent) the perspectives of people with lived experience. This evidence should be considered when selecting outcome measures for future studies.

As in all areas of research, there are opportunities for future studies to maximise the value of their data for evidence synthesis through more complete reporting. Reporting is crucial for the use and re-use of evidence from primary studies by decision-makers and in evidence synthesis. Primary studies are often used in evidence synthesis to address questions that differ from those addressed by the study investigators. For example, many of the studies screened for inclusion in this systematic review aimed to examine effects of GAH but included data for those who received puberty suppression that was not reported separately. This data could have contributed to the synthesis had it been reported separately. The ease with which supplementary information can be made available electronically substantially reduces barriers to complete reporting.

### 6.2.3 Equity related implications for research

There is a need for more complete reporting of equity-related characteristics, to enable greater understanding of intersecting factors that may exacerbate experiences of inequity (as per PROGRESS plus (54)). Growing awareness of the need to understand equity implications may mean that future studies address this gap. Considering equity-related guidance from Cochrane, GRADE and the Guideline International Network may help investigators plan for optimal reporting of equity-related data in primary studies(54, 115-117). In interpreting evidence, care needs to be taken to ensure that factors that limit access to treatment for certain groups (such as place of residence, language barriers, cultural safety or unconscious bias) are not conflated with a lack of treatment effectiveness.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

# 7 Additional information

## 7.1 Acknowledgements

We are grateful to the people with lived experience, members of the Committee and members of the Reference Panel who have generously shared their expertise and experience to ensure the relevance of this review to those affected by the decisions that the review will inform.

We thank the Gender Guidelines team at the ONHMRC with whom we worked to develop the protocol for this review. The team provided essential advice and support needed to maximise the relevance and usability of the review for decision-making. This included liaising with members of the Committee and Reference Panel to arrange targeted consultation during the planning and conduct of the review, coordinating the outcome prioritisation process, writing the background text on the Gender Guidelines, and coordinating feedback at all stages.

Professor Joanne McKenzie was supported by an Australian National Health and Medical Research Council (NHMRC) Investigator Grant (GNT2009612). Associate Professor Matthew J Page was supported by an NHMRC Investigator Grant (GNT2033917).

## 7.2 Contributions of authors

| | |
|---|---|
| Sue E Brennan | Led the design and writing of the protocol in collaboration with SM and JM. Responsible for leading the overall review, development of the data collection system, screening and triage of studies, extraction of study characteristics, verification of risk of bias assessments, management of study data, assessment of certainty of the evidence, coordination of team member contributions, and collation of results. Prepared the final report (with contributions from co-authors as indicated). |
| Steve McDonald (information specialist) | Designed and developed the search strategy and wrote the search and screening methods for the protocol. Responsible for leading the search and selection of studies. Conducted the search. Screened records and full text retrieved from the search. Collated screening decisions for reporting. Revised the methods sections of the report as per protocol. Wrote the results for the search and screening in report. |
| Simon Turner (biostatistician) | Extracted results data from primary studies, conducted data checking and comparison to ensure concordance of extracted data, prepared data for synthesis and reporting, conducted the analysis, prepared forest plots, drafted the Supplementary Appendix S1.C that reports on the preparation of data for meta-analysis. |

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

| | |
|---|---|
| Matthew J Page<br><br>(review methodologist) | Piloted the risk of bias assessment. Developed standardised wording to report assessments. Conducted and reported assessment of all studies. |
| David Clark<br><br>(biostatistician) | Contributed to the development of the REDcap database for triaging studies and extracting data. Developed REDcap form for ROBINS-I assessment and piloted the forms. Extracted results data from primary study. |
| Joanne E McKenzie<br><br>(senior biostatistician) | Designed and wrote sections of the protocol including: study design eligibility, measures of treatment effect, dealing with missing data, assessment of heterogeneity, unit of analysis issues, and data synthesis. Responsible for leading the analysis. Piloted the risk of bias assessment and co-developed standardised wording to report assessments, and piloted quantitative data extraction. Oversaw the collection of results data, preparation for meta-analysis of this data, and implementation of analysis plan. Checked certainty assessments. Revised the methods sections of the report as per protocol. |

## 7.3 Declarations of interest

As per declarations made by each author to the NHMRC, all authors declared that to the best of their knowledge they did not have any pecuniary, professional or personal interests that would constitute a conflict of interest.

## 7.4 Sources of support

### Internal sources

The facilities and resources used in producing this review (including access to databases, journal articles, software) were provided through Monash University and the School of Public Health and Preventive Medicine.

### External sources

This review is funded by the Australian Government Department of Health and Aged Care via the NHMRC, under Official Order 2025-26C001. While the evidence team worked in close collaboration with the ONHMRC to ensure the relevance of the planned review, the design of the methods, conduct and reporting of the review was done independently by the evidence team (i.e. the authors of the protocol and review report).

## 7.5 Registration and protocol

This review was not registered on PROSPERO. The draft protocol for the review underwent methodological review in March 2026 by an independent reviewer contracted by the NHMRC. The methodological review was conducted

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License


FOR PUBLIC CONSULTATION

prior to formal title and abstract screening. The draft was also reviewed by the Committee, the Reference Panel, and the ONHMRC (see governance structure at https://www.nhmrc.gov.au/health-advice/guidelines-care-trans-and-gender-diverse-people/governance). Any changes to the methodology based on Committee or Reference Panel feedback (e.g. to eligibility criteria) were finalised before screening of studies commenced. The revisions are documented in the Supplement 1, Appendix A1.E. Differences between the review and protocol (i.e. material changes to the methods reported in the protocol) are reported in the Supplement 1, Appendix A1.D.

## Data, code and other materials [availability of]

Data, code and other materials relating to this review will be made available on an open access repository on completion of the project for which this review is being conducted.

## Disclosure of artificial intelligence use

Artificial intelligence was not used to draft, write or edit text in the protocol or review report. It was used to provide initial leads on background sources related to defining and measuring adverse events and bone health, augmenting targeted consultation with members of the Committee and Reference Panel. Any guidance and research papers identified for this purpose using AI were retrieved, read and if relevant the key points summarised by one of the review authors (SB). AI was not used to summarise or interpret the identified sources, and subsequent leads were identified manually through the reference lists of retrieved literature.

## Positionality statement

We acknowledge that our worldview, professional and lived experience influence the design and reporting of this review. We are research methodologists who specialise in evidence synthesis. Our leads (SB, SM, JM) have each worked in the field of evidence synthesis for more than two decades. Our primary training is in the fields of science and education (SB), mathematics and statistics (JM, ST), social and political sciences (SM) and psychology (MP). Five of us completed doctoral research in speciality fields of evidence synthesis (complex reviews, information science, bias, biostatistics and meta-analysis). We undertake research on the methods used for evidence synthesis and contribute to international guidance on the conduct of systematic reviews (e.g. Cochrane Handbook, GRADE book). Four of us authored the PRISMA statement for reporting reviews, two led the statement (MP, JM). One of us co-authored the health equity extension for the Guidelines International Network checklist extension (SB).

We have undertaken evidence syntheses across diverse topics but have not previously contributed to research directly related to the health or care of people who are trans or gender diverse. Across all our reviews, we are committed to transparent and fair evidence assessment, that empowers individuals to make informed, values-based decisions about their health and wellbeing. This means we strive to address the outcomes that matter to people directly affected by the interventions we evaluate, and we explicitly consider health equity to ensure that the evidence can be interpreted appropriately for groups that experience inequity.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

# 8 References


1. Tugwell P, Welch VA, Karunananthan S, Maxwell LJ, Akl EA, Avey MT, et al. When to replicate systematic reviews of interventions: consensus checklist. BMJ. 2020;370:m2864.
2. Power J, Kauer S, Fisher C, Bellamy R, Bourne A. 7th National Survey of Australian Secondary Students and Sexual Health 2021 (ARCSHS Monograph Series No. 133). Australian Research Centre in Sex, Health & Society, La Trobe University; 2022.
3. Fisher C, Waling A, Kerr L, Bellamy R, Ezer P, Mikolajczak M, et al. 6th National Survey of Australian Secondary Students and Sexual Health 2018 (ARCSHS Monograph Series No. 113). Australian Research Centre in Sex, Health & Society, La Trobe University; 2019.
4. Australian Human Rights Commission. Equal identities: A human rights review of the experiences of trans and gender diverse people in Australia. Sydney, New South Wales; 2026 March 2026.
5. Hill AO LA, Jones J, McGowan I, Carman M, Parsons M, Power J, Bourne A Writing Themselves In 4: The health and wellbeing of LGBTQA+ young people in Australia. National report, monograph series number 124. Melbourne: Australian Research Centre in Sex, Health and Society, La Trobe University.; 2021.
6. Marino JL, Werner-Seidler A, Maston K, Lin A, Perry Y, Bista S, et al. Sexuality and Gender Diversity Among Adolescents in Australia, 2019-2021. JAMA Network Open. 2024;7(10):e2444187–e.
7. Strauss P, Cook, A., Winter, S., Watson, V., Wright Toussaint, D., Lin, A. Trans Pathways: the mental health experiences and care pathways of trans young people. Summary of results. Telethon Kids Institute, Perth, Australia.; 2017.
8. Telfer MM, Tollit MA, Pace CC, Pang KC. Australian standards of care and treatment guidelines for transgender and gender diverse children and adolescents. Med J Aust. 2018;209(3):132–6.
9. Coyne CA, Yuodsnukis BT, Chen D. Gender Dysphoria: Optimizing Healthcare for Transgender and Gender Diverse Youth with a Multidisciplinary Approach. Neuropsychiatr Dis Treat. 2023;19:479–93.
10. Ogarrio K, Durbin-Matrone R, De Leon B, Wright LA, Restar A, Singh AA, et al. Forms of transphobia and their influence on health outcomes among transgender, nonbinary, and gender diverse individuals: a global systematic review. International Journal of Transgender Health. 2025:1–34.
11. Bailey S, Newton N, Perry Y, Davies C, Lin A, Marino JL, et al. Minority stressors, traumatic events, and associations with mental health and school climate among gender and sexuality diverse young people in Australia: Findings from a nationally representative cohort study. Journal of Adolescence. 2024;96(2):275–90.
12. Coyne CA, Poquiz JL, Janssen A, Chen D. Evidence-Based Psychological Practice for Transgender and Non-Binary Youth: Defining the Need, Framework for Treatment Adaptation, and Future Directions. Evidence-Based Practice in Child and Adolescent Mental Health. 2020;5(3):340–53.
13. Weixel T, Whitehead J, Chen D. Recent Findings on Psychosocial Outcomes of Gender-Affirming Medical Care for Transgender Youth. The Journal of Clinical Endocrinology & Metabolism. 2025;111(2):e640–e5.
14. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.)2013.
15. Galupo MP, Lindley L, Clements ZA. An integrated framework for conceptualizing and measuring gender dysphoria: Development and psychometric evaluation of the Multidimensional Gender Dysphoria Measure. Am Psychol. 2025.
16. Horton C, Pearce R, Veale JF, Kerley L, Pang KC. Pathologising Legacy Measures and over-assessment: reconceptualising ethical assessment and data collection practices in trans children's healthcare. International Journal of Transgender Health. 2026:1–19.
17. O'Connell MA, Nguyen TP, Ahler A, Skinner SR, Pang KC. Approach to the Patient: Pharmacological Management of Trans and Gender-Diverse Adolescents. J Clin Endocrinol Metab. 2022;107(1):241–57.


FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

18. Olson-Kennedy J, Chan YM, Garofalo R, Spack N, Chen D, Clark L, et al. Impact of Early Medical Treatment for Transgender Youth: Protocol for the Longitudinal, Observational Trans Youth Care Study. JMIR Res Protoc. 2019;8(7):e14434.

19. Wright D, Pang KC, Giordano S, Gillam L. Evaluating the benefits and risks of puberty blockers and gender-affirming hormones for transgender adolescents. J Paediatr Child Health. 2025;61(1):7–11.

20. Council of Australian Therapeutic Advisory Groups. Rethinking medicines decision-making in Australian Hospitals. Guiding principles for the quality use of off-label medicine. 2013.

21. Gazarian M, Kelly M, McPhee JR, Graudins LV, Ward RL, Campbell TJ. Off-label use of medicines: consensus recommendations for evaluating appropriateness. Med J Aust. 2006;185(10):544–8.

22. Gray SG, McGuire TM. Navigating off-label and unlicensed medicines use in obstetric and paediatric clinical practice. Journal of Pharmacy Practice and Research. 2019;49(4):389–95.

23. New Zealand Ministry of Health. Impact of Puberty Blockers in Gender-Dysphoric Adolescents: An evidence brief. 2024.

24. New Zealand Ministry of Health. Addendum to Impact of Puberty Blockers in Gender-Dysphoric Adolescents: An evidence brief. 2024.

25. Bragge N, Delafosse G, Temple-Smith S. Evidence for effective interventions for children and young people with gender dysphoria—update. An Evidence Check rapid review brokered by the Sax Institute (www saxinstitute org au) for the NSW Ministry of Health. 2024.

26. Dopp AR, Peipert A, Buss J, De Jesús-Romero R, Palmer K, Lorenzo-Luaces L. Interventions for Gender Dysphoria and Related Health Problems in Transgender and Gender-Expansive Youth: A Systematic Review of Benefits and Risks to Inform Practice, Policy, and Research. Rand Health Q. 2025;12(2):2.

27. LaFleur J, Heath G, Gonzales V, Luloh M, Hankels R, Vadipour A, et al. Gender-affirming medical treatments for pediatric patients with gender dysphoria. University of Utah College of Pharmacy. 2024.

28. Ludvigsson JF, Adolfsson J, Hoistad M, Rydelius PA, Kristrom B, Landen M. A systematic review of hormone treatment for children with gender dysphoria and recommendations for research. Acta Paediatr. 2023;112(11):2279–92.

29. Miroshnychenko A, Roldan Y, Ibrahim S, Kulatunga-Moruzi C, Montante S, Couban R, et al. Puberty blockers for gender dysphoria in youth: A systematic review and meta-analysis. Arch Dis Child. 2025;110(6):429–36.

30. Porritt K, Whitehorn A, Stephenson M. Independent Review of Evidence-Base for the Use of Puberty Blockers and Gender Affirming Hormone Therapy for Children and Adolescents. Independent Review Advice Report for Queensland Health. 2025.

31. Rew L, Young CC, Monge M, Bogucka R. Review: Puberty blockers for transgender and gender diverse youth-a critical review of the literature. Child Adolesc Ment Health. 2021;26(1):3–14.

32. Taylor J, Mitchell A, Hall R, Heathcote C, Langton T, Fraser L, et al. Interventions to suppress puberty in adolescents experiencing gender dysphoria or incongruence: a systematic review. Arch Dis Child. 2024;109(Suppl 2):s33–s47.

33. Thompson L, Sarovic D, Wilson P, Irwin L, Visnitchi D, Samfjord A, et al. A PRISMA systematic review of adolescent gender dysphoria literature: 3) treatment. PLOS Glob Public Health. 2023;3(8):e0001478.

34. Tornese G, Di Mase R, Munarin J, Ciancia S, Santamaria F, Fava D, et al. Use of gonadotropin-releasing hormone agonists in transgender and gender diverse youth: a systematic review. Front Endocrinol (Lausanne). 2025;16:1555186.

35. McKenzie JE, Brennan SE. Overviews of systematic reviews: great promise, greater challenge. Systematic reviews. 2017;6(1):185.

36. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024). Cochrane; 2024. Available from: www.cochrane.org/handbook.

37. 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.

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License



FOR PUBLIC CONSULTATION

38. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021;372:n160.

39. Bloom TM, Nguyen TP, Lami F, Pace CC, Poulakis Z, Telfer M, et al. Measurement tools for gender identity, gender expression, and gender dysphoria in transgender and gender-diverse children and adolescents: a systematic review. Lancet Child Adolesc Health. 2021;5(8):582–8.

40. Reeves BC, Deeks JJ, Higgins JPT, Shea B, Tugwell P, Wells GA. Chapter 24: Including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 65 (updated August 2024) Cochrane, 2024 Available from: http://wwwcochraneorg/handbook.

41. Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels. J Clin Epidemiol. 2017;89:30–42.

42. Higgins JPT, Savovic J, Page MJ, Sterne JAC. Revised Cochrane risk-of-bias tool for randomized trials (RoB 2). Available from https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2. Access date: 6 February 2026.

43. Sterne JAC, Hernán MA, McAleenan A, Reeves BC, Higgins JPT. Chapter 25: Assessing risk of bias in a non-randomized study. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 65 (updated August 2024) Cochrane, 2024 Available from: http://wwwcochraneorg/handbook.

44. Hernan MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. JAMA. 2022;328(24):2446–7.

45. WHO. International Classification of Diseases Eleventh Revision (ICD-11). Geneva, Switzerland: World Health Organization; 2022.

46. McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnson RV, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Welch V, editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated August 2023) Cochrane, 2024 Available from wwwtrainingcochraneorg/handbook. 2nd ed. Chichester (UK): John Wiley & Sons; 2024.

47. McCaul M, Naude C, Neumann I, Brennan S, Meerpohl K, Davoli M, et al. Outcomes. The GRADE Book version 1.0 (updated December 2025). The GRADE Working Group. Available from: https://book.gradepro.org.2025.

48. Moore JK, Rayner C, Skinner SR, Wynne K, Cavve BS, Fraser B, et al. Cass Review does not guide care for trans young people. Med J Aust. 2025;223(7):331–7.

49. Thomas J, Kneale D, McKenzie J, Brennan S, Bhaumik S. Chapter 2: Determining the scope of the review and the questions it will address. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Welch V, editors. Cochrane Handbook for Systematic Reviews of Interventions version 65 (updated August 2023) Cochrane, 2024 Available from wwwtrainingcochraneorg/handbook. 2nd ed. Chichester (UK): John Wiley & Sons; 2024.

50. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol. 2016;75:40–6.

51. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019;95:103208.

52. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377–81.

53. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014;348:g1687.

54. Welch VA, Petkovic J, Jull J, Hartling L, Klassen T, Kristjansson E, et al. Chapter 16: Equity and specific populations [last updated October 2019]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors.


© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License



FOR PUBLIC CONSULTATION

Cochrane Handbook for Systematic Reviews of Interventions version 65 (updated August 2024) Cochrane, 2024 Available from: http://wwwcochraneorg/handbook.

55. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.

56. Sterne JAC, Higgins JPT, Elbers RG, Reeves BC. Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I V2): detailed guidance, updated November 2025. Available from www.riskofbias.info [accessed 6 February 2026].

57. Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. J Clin Epidemiol. 2019;111:105–14.

58. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med. 2002;21(11):1575–600.

59. Zeng L, Brignardello-Petersen R, Hultcrantz M, Siemieniuk RAC, Santesso N, Traversy G, et al. GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings. J Clin Epidemiol. 2021;137:163–75.

60. Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ. 2020;369:m1714.

61. Brennan SE, Johnston RV. Research Note: Interpreting findings of a systematic review using GRADE methods. J Physiother. 2023;69(3):198–202.

62. Zeng L, Brignardello-Petersen R, Hultcrantz M, Mustafa RA, Murad MH, Iorio A, et al. GRADE Guidance 34: update on rating imprecision using a minimally contextualized approach. J Clin Epidemiol. 2022;150:216–24.

63. Balk EM, Earley A, Patel K, Trikalinos TA, Dahabreh IJ. AHRQ Methods for Effective Health Care. Empirical Assessment of Within-Arm Correlation Imputation in Trials of Continuous Outcomes. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012.

64. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539–58.

65. Page MJ, Higgins JPT, Sterne JAC. Chapter 13: Assessing risk of bias due to missing results in a synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 65 (updated August 2024) Cochrane, 2024 Available from: http://wwwcochraneorg/handbook.

66. Carmichael P, Butler G, Masic U, Cole TJ, De Stavola BL, Davidson S, et al. Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK. PLoS One. 2021;16(2):e0243894.

67. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol. 2008;61(10):991–6.

68. Wilkinson J, Heal C, Flemyng E, Antoniou GA, Aburrow T, Alfirevic Z, et al. INSPECT-SR: a tool for assessing trustworthiness of randomised controlled trials. medRxiv. 2025.

69. Deeks JJ, Higgins JPT, Altman DG, McKenzie JE, Veroniki A. Chapter 10: Analysing data and undertaking meta-analyses [last updated November 2024]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 65 (updated August 2024) Available from http://wwwcochraneorg/handbook: Cochrane; 2024.

70. van de Grift TC, van Gelder ZJ, Mullender MG, Steensma TD, de Vries ALC, Bouman MB. Timing of Puberty Suppression and Surgical Options for Transgender Youth. Pediatrics. 2020;146(5).

71. Neumann I, Schünemann H. The GRADE Book version 1.0 (updated December 2025). The GRADE Working Group. Available from: https://book.gradepro.org.

72. Schünemann HJ, Higgins JPT, Vist GE, Glasziou PP, Akl EA, Skoetz N, et al. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T,

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License



FOR PUBLIC CONSULTATION

Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 65 (updated August 2024) Cochrane, 2024 Available from: http://wwwcochraneorg/handbook.

73. Goldkuhle M, Kreuzberger N, Neumann I, Brennan SE, Meerpohl J, Langendam M, et al. Principles for assessing the certainty of interventions. In: Neumann I, Schünemann H (Editors). The GRADE Book version 1.0 (updated Aug 2025. The GRADE Working Group. Available from: https://book.gradepro.org/guideline/principles-for-assessing-the-certainty-of-interventions. 2026.

74. Neumann I, Morgan R, Langendam M, Brozek J, Cuellos C, Brennan SE, et al. Limitations in the design or execution of randomized trials. In: Neumann I, Schünemann H (Editors). The GRADE Book version 1.0 (updated Aug 2025. The GRADE Working Group. Available from: https://book.gradepro.org/guideline/risk-of-bias-randomized-trials. 2026.

75. Neumann I, Sousa-Pinto B, Meerpohl J, Dahm P, Brennan SE, Alonso Coello P, et al. Inconsistency. In: Neumann I, Schünemann H (Editors). The GRADE Book version 1.0 (updated July 2025. The GRADE Working Group. Available from: https://book.gradepro.org/guideline/inconsistency. 2026.

76. Guyatt G, Zhao Y, Mayer M, Briel M, Mustafa R, Izcovich A, et al. GRADE Guidance 36: Updates to GRADE's approach to addressing inconsistency. J Clin Epidemiol. 2023;158:70–83.

77. Murad MH, Neumann I, Brozek J, Langendam M, Dahm P, Schünemann HJ. Imprecision. In: Neumann I, Schünemann H (Editors). The GRADE Book version 1.0 (updated Aug 2025. The GRADE Working Group. Available from: https://book.gradepro.org/guideline/imprecision. 2025.

78. Schünemann HJ, Neumann I, Brennan SE, Meerpohl J, Langendam M, Davoli M, et al. Indirectness. In: Neumann I, Schünemann H (Editors). The GRADE Book version 1.0 (updated May 2026. The GRADE Working Group. Available from: https://book.gradepro.org/guideline/indirectness. 2026.

79. Toews I, Bruschettini M, Akl E, Skoetz N, Langendam M, Alonso Coello P, et al. Dissemination bias. In: Neumann I, Schünemann H (Editors). The GRADE Book version 1.0 (updated Aug 2025. The GRADE Working Group. Available from: https://book.gradepro.org/guideline/dissemination-bias. 2025.

80. Schunemann HJ, Brozek J, Guyatt G, Oxman AD, editors. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Hamilton, Canada: McMaster University; 2013.

81. Piggott T, Morgan RL, Cuello-Garcia CA, Santesso N, Mustafa RA, Meerpohl JJ, et al. Grading of Recommendations Assessment, Development, and Evaluations (GRADE) notes: extremely serious, GRADE's terminology for rating down by three levels. J Clin Epidemiol. 2020;120:116–20.

82. Santesso N, Carrasco-Labra A, Langendam M, Brignardello-Petersen R, Mustafa RA, Heus P, et al. Improving GRADE Evidence Tables part 3: Guidance for useful GRADE certainty in the evidence judgments through explanatory footnotes. J Clin Epidemiol. 2016;74:28–39.

83. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, et al. GRADE guidelines: 26. informative statements to communicate the findings of systematic reviews of interventions. J Clin Epidemiol. 2019;119:126–35.

84. Cuello-Garcia CA, Morgan RL, Santesso N, Alonso-Coello P, Brignardello-Petersen R, Schwingshackl L, et al. GRADE Guidance 44: strategies to enhance the utilization of randomized and non-randomized studies in evidence syntheses of health interventions. J Clin Epidemiol. 2026;190:112086.

85. Costa R, Dunsford M, Skagerberg E, Holt V, Carmichael P, Colizzi M. Psychological Support, Puberty Suppression, and Psychosocial Functioning in Adolescents with Gender Dysphoria. J Sex Med. 2015;12(11):2206–14.

86. de Vries AL, McGuire JK, Steensma TD, Wagenaar EC, Doreleijers TA, Cohen-Kettenis PT. Young adult psychological outcome after puberty suppression and gender reassignment. Pediatrics. 2014;134(4):696–704.

87. de Vries AL, Steensma TD, Doreleijers TA, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. J Sex Med. 2011;8(8):2276–83.

88. Fisher AD, Ristori J, Romani A, Cassioli E, Mazzoli F, Cocchetti C, et al. Back to the Future: Is GnRHa Treatment in Transgender and Gender Diverse Adolescents Only an Extended Evaluation Phase? J Clin Endocrinol Metab. 2024;109(6):1565–79.

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License



FOR PUBLIC CONSULTATION

89. Karamanis G, Frisell T, Holmberg M, Halldin M, Sylvén S, Skalkidou A, et al. Incidence of Idiopathic Intracranial Hypertension in Individuals With Gonadotropin-Releasing Hormone Analogue Treatment for Gender Dysphoria in Sweden. JAMA Pediatr. 2023;177(7):726–7.

90. Lahaije FAH, van Setten PA, Levels W, Becking-Malpasso K, Claahsen-van der Grinten HL. GnRH analogs as a monotherapy in transgender and gender-diverse adolescents: clinical insights from a single-center study. Endocrine connections. 2025;14(8).

91. Lavender R, Shaw S, Maninger JK, Butler G, Carruthers P, Carmichael P, et al. Impact of Hormone Treatment on Psychosocial Functioning in Gender-Diverse Young People. LGBT Health. 2023;10(5):382–90.

92. McGregor K, McKenna JL, Williams CR, Barrera EP, Boskey ER. Association of Pubertal Blockade at Tanner 2/3 With Psychosocial Benefits in Transgender and Gender Diverse Youth at Hormone Readiness Assessment. J Adolesc Health. 2024;74(4):801–7.

93. McPherson S, Freedman DEP. Psychological Outcomes of 12-15-Year-Olds with Gender Dysphoria Receiving Pubertal Suppression in the UK: Assessing Reliable and Clinically Significant Change. J Sex Marital Ther. 2024;50(3):315–25.

94. Nunes-Moreno M, Furniss A, Cortez S, Davis SM, Dowshen N, Kazak AE, et al. Mental Health Diagnoses and Suicidality Among Transgender Youth in Hospital Settings. LGBT Health. 2025;12(1):20–8.

95. Schagen SE, Cohen-Kettenis PT, Delemarre-van de Waal HA, Hannema SE. Efficacy and Safety of Gonadotropin-Releasing Hormone Agonist Treatment to Suppress Puberty in Gender Dysphoric Adolescents. J Sex Med. 2016;13(7):1125–32.

96. Turban JL, King D, Carswell J, Keuroghlian AS. 6.29 Treatment with Pubertal Suppression during Adolescence Is Associated with Lower Odds of Lifetime and Past-Year Suicidal Ideation among Transgender Adults in the United States. J Am Acad Child Adolesc Psychiatry. 2019;58(10 Supplement):S280.

97. Turban JL, King D, Carswell JM, Keuroghlian AS. Pubertal Suppression for Transgender Youth and Risk of Suicidal Ideation. Pediatrics. 2020;145(2).

98. van der Miesen AIR, Steensma TD, de Vries ALC, Bos H, Popma A. Psychological Functioning in Transgender Adolescents Before and After Gender-Affirmative Care Compared With Cisgender General Population Peers. J Adolesc Health. 2020;66(6):699–704.

99. Zaitoon H, Kapusta L, Israeli G, Segev-Becker A, Beer G, Lowenthal A, et al. QT interval changes in transgender/gender diverse adolescents receiving gonadotropin-releasing hormone agonist and gender-affirming hormonal treatment. Int J Transgend Health. 2026;27(2):1033–44.

100. Becker-Hebly I, Fahrenkrug S, Campion F, Richter-Appelt H, Schulte-Markwort M, Barkmann C. Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: a descriptive study from the Hamburg Gender Identity Service. Eur Child Adolesc Psychiatry. 2021;30(11):1755–67.

101. Navabi B, Tang K, Khatchadourian K, Lawson ML. Pubertal Suppression, Bone Mass, and Body Composition in Youth With Gender Dysphoria. Pediatrics. 2021;148(4).

102. Schagen SEE, Wouters FM, Cohen-Kettenis PT, Gooren LJ, Hannema SE. Bone Development in Transgender Adolescents Treated With GnRH Analogues and Subsequent Gender-Affirming Hormones. J Clin Endocrinol Metab. 2020;105(12):e4252–63.

103. Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry. 2010;49(10):980–9.

104. Robson EM, Husin HM, Ghazaleh Dashti S, Vijayakumar N, Moreno-Betancur M, Moran P, et al. Tracking the course of depressive and anxiety symptoms across adolescence (the CATS study): a population-based cohort study in Australia. Lancet Psychiatry. 2025;12(1):44–53.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

105. Tienforti D, Marinelli L, Terrana G, Di Geronimo R, Spagnolo L, Vervalcke J, et al. Bone Accrual During Puberty Suppression and Gender-Affirming Therapy in Transgender Adolescents: A Systematic Review and Meta-Analysis. JAMA Pediatr. 2026.

106. Neumann I, Alonso-Coello P, Vandvik PO, Agoritsas T, Mas G, Akl EA, et al. Do clinicians want recommendations? A multicenter study comparing evidence summaries with and without GRADE recommendations. Journal of Clinical Epidemiology. 2018;99:33–40.

107. Neumann I, Brennan S, Meerpohl J, Davoli M, Alonso Coello P, Akl A, et al. Overview of the GRADE approach. The GRADE Book version 1.0 (updated May 2026). The GRADE Working Group. Available from: https://book.gradepro.org.

108. Cuello-Garcia CA, Santesso N, Morgan RL, Verbeek J, Thayer K, Ansari MT, et al. GRADE guidance 24 optimizing the integration of randomized and non-randomized studies of interventions in evidence syntheses and health guidelines. J Clin Epidemiol. 2022;142:200–8.

109. McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 65 (updated August 2024) Cochrane, 2024 Available from: http://wwwcochraneorg/handbook.

110. International Society for Clinical Densitometry. Pediatric positions. ISCD. https://iscd.org/learn/official-positions/pediatric-positions/ 2019 [Available from: https://iscd.org/learn/official-positions/pediatric-positions/.

111. Doyle DM, de Vries ALC. Systematic reviews and the needs of the adolescent transgender health care field. The Lancet. 2026;408(10557):861–6.

112. Peryer G, Golder S, Junqueira D, Vohra S, Loke YK. Chapter 19: Adverse effects. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 65 Cochrane, 2024 Available from cochraneorg/handbook2019.

113. Dahabreh IJ, Bibbins-Domingo K. Causal Inference About the Effects of Interventions From Observational Studies in Medical Journals. JAMA. 2024;331(21):1845–53.

114. Wijesuriya R, Buckingham P, Moreno-Betancur M, Ghazaleh D, Becker D, Ravipati T, et al. Comparative effectiveness of gender affirming hormones on body image, gender dysphoria, trans pride, mental health and quality of life in trans and gender diverse adolescents: a target trial emulation (statistical analysis protocol). Figshare [Internet]. 2025. Available from: https://doi.org/10.25374/MCRI.30724058.v1.

115. Dewidar O, Darzi AJ, Sayfi S, Pardo JP, Welch V, Wright GC, et al. Seven principles for integrating health equity considerations in the practice guideline enterprise. J Clin Epidemiol. 2025;182:111777.

116. Dewidar O, Sayfi S, Pardo JP, Welch V, Wright GC, Akl EA, et al. Enhancing health equity considerations in guidelines: health equity extension of the GIN-McMaster Guideline Development Checklist. eClinicalMedicine. 2025;82:103135.

117. Welch VA, Akl EA, Guyatt G, Pottie K, Eslava-Schmalbach J, Ansari MT, et al. GRADE equity guidelines 1: considering health equity in GRADE guideline development: introduction and rationale. J Clin Epidemiol. 2017;90:59–67.

118. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. Arch Dis Child. 1969;44(235):291–303.

119. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child. 1970;45(239):13–23.

120. Delemarre-van de Waal HA, Cohen-Kettenis PT. Clinical management of gender identity disorder in adolescents: a protocol on psychological and paediatric endocrinology aspectsThis paper was presented at the 4th Ferring Pharmaceuticals International Paediatric Endocrinology Symposium, Paris (2006). Ferring Pharmaceuticals has supported the publication of these proceedings. European Journal of Endocrinology. 2006;155(Supplement_1):S131–S7.

121. Donaghy OJE, Cobham V, Lin A. Screening adolescent transgender-related distress: Gender preoccupation and stability questionnaire demonstrates excellent criterion validity with multi-disciplinary, pediatric gender specialist assessment. Int J Transgend Health. 2024;27(1):1–9.


© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License



FOR PUBLIC CONSULTATION

122. Hakeem A, Črnčec R, Asghari-Fard M, Harte F, Eapen V. Development and validation of a measure for assessing gender dysphoria in adults: The Gender Preoccupation and Stability Questionnaire. International Journal of Transgenderism. 2016;17(3-4):131–40.

123. McGuire JK, Berg D, Catalpa JM, Morrow QJ, Fish JN, Nic Rider G, et al. Utrecht Gender Dysphoria Scale - Gender Spectrum (UGDS-GS): Construct validity among transgender, nonbinary, and LGBQ samples. Int J Transgend Health. 2020;21(2):194–208.

124. Deogracias JJ, Johnson LL, Meyer-Bahlburg HF, Kessler SJ, Schober JM, Zucker KJ. The gender identity/gender dysphoria questionnaire for adolescents and adults. J Sex Res. 2007;44(4):370–9.

125. Clements Ph DZ, Pulice-Farrow Ph DL, Paz Galupo Ph DM. "Outdated and offensive": trans and nonbinary individuals' responses to clinical scales of gender dysphoria. J Sex Med. 2025;22(11):2159–60.

126. de Groot M, van Hoeken D, Keski-Rahkonen A. Eating Disorder Symptoms, Body Satisfaction, and Self-Esteem Among Dutch Mental Health Outpatient Youth With Gender Incongruence. Psychology of Sexual Orientation and Gender Diversity. 2026.

127. Lindgren TW, Pauly IB. A body image scale for evaluating transsexuals. Arch Sex Behav. 1975;4(6):639–56.

128. van de Grift TC, Cohen-Kettenis PT, Steensma TD, De Cuypere G, Richter-Appelt H, Haraldsen IR, et al. Body Satisfaction and Physical Appearance in Gender Dysphoria. Arch Sex Behav. 2016;45(3):575–85.

129. Institute NIoHNC. Common terminology criteria for adverse events (CTCAE) v6.0 2026 [Available from: https://dctd.cancer.gov/research/ctep-trials/for-sites/adverse-events.

130. WHO. The World Health Organization Quality of Life (WHOQOL): Measuring Quality of Life 2012 [Available from: https://www.who.int/tools/whoqol.

131. Anderson RT, Aaronson NK, Bullinger M, McBee WL. A review of the progress towards developing health-related quality-of-life instruments for international clinical studies and outcomes research. Pharmacoeconomics. 1996;10(4):336–55.

132. Karimi M, Brazier J. Health, health-related quality of life, and quality of life: what is the difference? Pharmacoeconomics. 2016;34(7):645–9.

133. Kennedy SL, Cornacchi SD, Kaur MN, Morrison S, Rae C, Johnson N, et al. Development of a Patient-Reported Outcome Measure for Youth Receiving Gender-Affirming Care: The GENDER-Q Youth Module. Transgend Health. 2025;10(4):375–85.

134. Testa RJ, Habarth J, Peta J, Balsam K, Bockting W. Development of the gender minority stress and resilience measure. Psychology of sexual orientation and gender diversity. 2015;2(1):65.

135. Sim L, Wang Z, Croarkin PE, McKean AJ, Mohammed KS, Rajjo TI, et al. Management of Suicidal Thoughts and Behaviors in Youth: A Systematic Review. Rockville (MD)2025.

136. Hetrick SE, McKenzie JE, Bailey AP, Sharma V, Moller CI, Badcock PB, et al. New generation antidepressants for depression in children and adolescents: a network meta-analysis. Cochrane Database Syst Rev. 2021;5(5):CD013674.

137. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry. 2011;168(12):1266–77.

138. Adams JE. Advances in bone imaging for osteoporosis. Nature Reviews Endocrinology. 2013;9(1):28–42.

139. Crabtree NJ, Arabi A, Bachrach LK, Fewtrell M, El-Hajj Fuleihan G, Kecskemethy HH, et al. Dual-Energy X-Ray Absorptiometry Interpretation and Reporting in Children and Adolescents: The Revised 2013 ISCD Pediatric Official Positions. Journal of Clinical Densitometry. 2014;17(2):225–42.

140. Weaver CM, Gordon CM, Janz KF, Kalkwarf HJ, Lappe JM, Lewis R, et al. The National Osteoporosis Foundation's position statement on peak bone mass development and lifestyle factors: a systematic review and implementation recommendations. Osteoporos Int. 2016;27(4):1281–386.

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License



FOR PUBLIC CONSULTATION

141. Ward LM, Weber DR, Munns CF, Högler W, Zemel BS. A Contemporary View of the Definition and Diagnosis of Osteoporosis in Children and Adolescents. J Clin Endocrinol Metab. 2020;105(5):e2088–97.

142. Grigoryan M, Guermazi A, Roemer FW, Delmas PD, Genant HK. Recognizing and reporting osteoporotic vertebral fractures. Eur Spine J. 2003;12 Suppl 2(Suppl 2):S104–12.

143. Clark EM, Ness AR, Bishop NJ, Tobias JH. Association between bone mass and fractures in children: a prospective cohort study. J Bone Miner Res. 2006;21(9):1489–95.

144. Pang KC, Peri AJS, Chung HE, Telfer M, Elder CV, Grover S, et al. Rates of Fertility Preservation Use Among Transgender Adolescents. JAMA Pediatrics. 2020;174(9):890–1.

145. Baines HK, Quinn GP, Chen D, Nahata L. Reproductive Health in Trans and Gender Diverse Patients: Fertility related knowledge, attitudes, and decision-making among gender diverse youth - A narrative review. Reproduction. 2024;168(2).

146. Cavve BS, Bickendorf X, Ball J, Saunders LA, Thomas CS, Strauss P, et al. Reidentification With Birth-Registered Sex in a Western Australian Pediatric Gender Clinic Cohort. JAMA Pediatr. 2024;178(5):446–53.

FOR PUBLIC CONSULTATION

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License

# Appendices

## Appendix 1a. Outcome domains prioritised for this review

| | Outcome domain | Specification of domain (definition, inclusions) | Example measurement methods | Primary goal of treatment | Value to those affected | Rationale |
|---|---|---|---|---|---|---|
| Critical outcomes | | | | | | |
| 1 | Pubertal development | Progression of physical changes induced by endogenous sex hormones. These secondary sex characteristics include voice deepening, hair growth, breast development, final height, body habitus, and menstruation.<br><br>**Exclusions**: secondary sex characteristics of the affirmed gender induced by GAH (i.e. breast size measured after GAH treatment in individuals assigned male at birth would be ineligible even if the study compared those who had received PS to those who had not, since this outcome pertains primarily to the effects of GAH not PS). | Tanner's sexual maturation scale (118, 119).<br><br>Measures of developmental stage secondary sex characteristics (testicular volume, penile length, breast development, menstruation). | Yes | High<br><br>(pausing physical changes, including irreversible changes) | The main goal of puberty suppression is to temporarily "halt the development of secondary sex characteristics associated with puberty" (19), "allowing a young person and their family time to consider whether to progress to [gender affirming] treatment" (11). Longer-term implications also require consideration. Many physical changes that occur during pubertal development are permanent (late-stage breast development, lowering of the voice, facial hair) (120). Puberty suppression may prevent unwanted changes and reduce future adverse consequences such as invasive surgery in adulthood (92).<br><br>**Timeframe for follow-up:** during treatment; after discontinuation of puberty suppressants (sex hormone production is expected to re-commence inducing pubertal development) |

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License.

| | Outcome domain | Specification of domain (definition, inclusions) | Example measurement methods | Primary goal of treatment | Value to those affected | Rationale |
|---|---|---|---|---|---|---|
| 2 | Gender dysphoria | "Distress that is associated with the individual's subjective experience of gender incongruence" (10).<br><br>A broad definition of gender dysphoria is used, rather than narrower diagnostic criteria. Concepts and operational definitions of gender dysphoria (for measurement) vary in the extent to which they have been based on clinical perspectives (e.g. DSM-5, some common measurement tools) versus integrating the lived experiences of people who are trans and gender diverse (7). Concepts grounded in lived experience consider the influence of exposure to minority stress ( stemming from stigma, discrimination, non-acceptance etc) on body dysphoria (12). | Includes clinical diagnosis using recognised criteria (e.g. DSM) and scales measuring severity.<br><br>• Gender preoccupation and stability questionnaire (GPSQ) (121, 122)<br>• Multidimensional gender dysphoria scale (15)<br><br>Other widely used scales<br><br>• Utrecht Gender Dysphoria Scale (UGDS) (123)<br>• Gender Identify / Gender dysphoria questionnaire for adolescents and adults (GIDYQ-AA) (124)<br><br>Measures need careful interpretation due conceptual changes, outdated terminology, omission of non-binary gender identities, pathologisation etc (16, 39, 125). | Yes – to stabilise (no physical change) | High<br><br>(stabilising distress) | Temporarily halting puberty may prevent distress (or exacerbation of distress) associated with onset or progression of puberty (19). These stabilising effects rather than improvement are anticipated because puberty suppressants do not induce gender affirming physical changes.<br><br>**Timeframe for follow-up:** during treatment; end of treatment period |
| 3 | Body image and satisfaction | Body image and satisfaction are defined as "subjective evaluations of one's physical appearance" (126). Young trans and gender diverse people commonly report negative subjective evaluations of their physical appearance relating to body characteristics that are not congruent with their gender identity, with increasing risk of | Body image and satisfaction scales and other measures that provide indicators of whether a young person's embodiment goals are achieved.<br><br>Body image scale (BIS) (127, 128)<br><br>• separate body area scales (e.g. social and hair items, muscularity and posture, hip region, chest region, genitals) | Yes – to stabilise (no physical change) | High<br><br>(achieving embodiment goals) | A temporary halt in puberty may prevent deterioration in a young person's body image and satisfaction associated with development of secondary sex characteristics incongruent with gender identity (19). These stabilising effects rather than improvement are anticipated because puberty suppressants do not induce gender affirming physical changes. |

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License.

| | Outcome domain | Specification of domain (definition, inclusions) | Example measurement methods | Primary goal of treatment | Value to those affected | Rationale |
|---|---|---|---|---|---|---|
| | | dissatisfaction as puberty progresses. | • separate versions for people assigned female at birth and people assigned male at birth. | | | **Timeframe for follow-up:** during treatment; end of treatment period |
| 4 | Serious adverse events (SAEs) | Any serious untoward outcome that occurs during, or after, the use of puberty suppressants, but not necessarily caused by the puberty suppressant[8],[9] (112). An AE is defined as serious if it "results in death, is life-threatening, requires hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability or incapacity"[2].<br><br>Eligible outcomes include *spontaneously reported events* (e.g. a symptom that emerges or worsens after treatment commences) and *outcomes prespecified by study investigators* (e.g. an AE of special interest for which drug regulators require active monitoring) (112). | Reports of<br>• any measure of an AE that meets the definition of serious<br>• composite measures of SAEs (e.g. total number participants with SAEs) (112)<br>• AEs of special interest<br>We will use the Common terminology criteria for adverse events (CTCAE) or similar to confirm severity as needed (129). AEs of special interest are not necessarily serious but may elucidate any safety concerns or signal the need for active monitoring.<br><br>Outcomes specified elsewhere in this table are considered separately but effects will be reported under SAEs if serious undesirable effects are observed (i.e. important harm). | Yes – avoidance of serious adverse effects | Yes – avoidance of SAEs | Considering evidence about the risk of serious adverse effects is critical for balanced decision-making.<br><br>**Timeframe for follow-up:** during treatment; short- and longer-term follow-up after treatment if effects can be isolated from subsequent medical care |

[8] Definition based on the Glossary of International Council for harmonisation terms and definitions (version 10) and the Cochrane Handbook of Interventions version 6.5. In these sources, an important distinction is made between an adverse 'event' which may or may not be related to treatment and an adverse 'effect' (or 'reaction'). The latter is established by evaluating whether there is evidence that the event is attributable to the treatment (i.e. a causal effect).

[9] Description from the National Health and Medical Research Council (2016). Guidance: Safety monitoring and reporting in clinical trials involving therapeutic goods. Canberra: National Health and Medical Research Council.

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License.

| | Outcome domain | Specification of domain (definition, inclusions) | Example measurement methods | Primary goal of treatment | Value to those affected | Rationale |
|---|---|---|---|---|---|---|
| Important outcomes | | | | | | |
| 5 | Quality of life (QoL) and wellbeing | As defined by the World Health Organisation (WHO) "an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns." (130) (131, 132)<br>Related concepts that are eligible for this domain:<br>• QoL (overall, may include psychological, emotional and social wellbeing)<br>• Wellbeing | Generic or population specific QoL measures including:<br>• Child health utility 9D<br>• GENDER-Q Youth Module (HR-QoL module) (133) | No – may stabilise but depends on other outcomes / care | High | Quality of life has been identified as a highly valued outcome by trans and gender diverse young people accessing gender affirming care (18). However, effects of puberty suppression on QoL are expected only after other outcomes stabilise and further care is provided to affirm gender (3).<br>**Timeframe for follow-up:** end of treatment period |
| 6 | Social functioning | Any of the following are eligible for this domain (with a hierarchy to be agreed):<br>• Social functioning (at school, outside school, at home)<br>• Social relationships<br>• Family functioning, communication, support<br>Excludes psychological function which is considered under the mental health domain. | Social functioning scales from HR-QoL measures or specific measures of social functioning, for example:<br>• GENDER-Q Youth Module (HR-QoL module - social function scale) (133)<br>• Gender Minority Stress and Resilience Measure (community connectedness and pride scales) (134)<br>Some measures of QoL address this domain. Where results for a social functioning scale are reported | No – may stabilise but depends on other outcomes / care | High | Social functioning has been identified as a highly valued outcome by trans and gender diverse young people accessing gender affirming care (18). However, effects of puberty suppression on functioning are expected only after other outcomes stabilise and further care is provided to affirm gender (3).<br>**Timeframe for follow-up:** end of treatment period |

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License.

| | Outcome domain | Specification of domain (definition, inclusions) | Example measurement methods | Primary goal of treatment | Value to those affected | Rationale |
|---|---|---|---|---|---|---|
| | | | separately, they will be included in the synthesis for 'social functioning'. | | | |
| 7 | Satisfaction with care | Subjective experiences of gender affirming care, particularly those related to puberty suppressants<br>• Young person's satisfaction with care<br>• Family satisfaction with care | General and service-specific measures of experience of care including:<br>• Validated measures used by services that provide gender affirming care<br>• GENDER-Q Youth Module – experience of care scale (133)<br>• Measures of experience of care related to puberty suppression, and whether care aligned with personal values, preferences and goals, will be prioritised. | No – but important indicator of whether care aligns with individual values, preferences, goals | High (alignment of care with individual values, preferences, goals) | Experience of care has been identified as a highly valued outcome by trans and gender diverse young people accessing gender affirming care (18). Experience of care is included to capture whether the care received (access to puberty suppression) aligns with expressed values, preferences, and embodiment goals. (3).<br>**Timeframe for follow-up:** end of treatment period |
| 8 | Mental health | Any of the following would be eligible for this domain (with a hierarchy to be agreed):<br>• depression symptoms<br>• anxiety symptoms<br>• suicidal thoughts (ideation, severity of ideation and intent) and behaviours (e.g. attempts, self-harm with suicidal intent) (135).<br>Outcomes listed here relate to co-occurring mental health conditions (depression, anxiety and suicidality) or symptoms. | Includes clinical diagnosis using recognised criteria (e.g. DSM or ICD) and scales used to measure symptom severity (135, 136).<br>• Beck Depression Inventory (BDI)/Children's Depression Inventory (CDI)<br>• Spence children anxiety scales (self-report, parent)<br>• Patient Health Questionnaire (PHQ)-9<br>• Columbia-Suicide Severity Rating Scale (C-SSRS) (137)<br>• Ask Suicide-Screening Questions (ASQ) | No | High | While care for trans and gender diverse youth with co-occurring mental health symptoms or conditions is paramount, puberty suppressants are not a form of psychological support or mental health treatments. Any effects are expected to be dependent on stabilising other outcomes and provision of mental health care (12, 13). For this reason, mental health outcomes are included as important but not critical for decision-making.<br>**Timeframe for follow-up:** end of treatment period |

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License.

| | Outcome domain | Specification of domain (definition, inclusions) | Example measurement methods | Primary goal of treatment | Value to those affected | Rationale |
|---|---|---|---|---|---|---|
| | | | • Hospitalisations for suicidal thoughts or behaviours<br>• Emergency department visits for suicidal thoughts or behaviours | | | |
| 9 | Mental health – suicide | Death by suicide. This outcome will be considered separately from other mental health related outcomes because of severity, and the type and availability of data. | Medical record or direct inquiry with appropriate contact person at follow-up (136). | No | High | As per 'Mental health' |
| 10 | Adverse events of special interest: bone health | Fractures or functional outcomes related to bone health.<br>Surrogate measures of bone health, (e.g. BMC, aBMD, Z-scores), were not prioritised for this review[10] (46, 47). Bone strength is strongly associated with fracture risk but cannot be measured directly (138-140). DXA (dual-energy X-ray absorptiometry) scans provide proxy measures of bone strength; primarily bone mass (bone mineral content, BMC) and areal bone mineral density (aBMD) (138-141).[11] | Any report of fracture (e.g. count, rate) or functional impact including new vertebral fractures (e.g. non-traumatic compression fractures confirmed by imaging)<br>• new non-vertebral fractures (e.g. asymptomatic low trauma [fragility] fractures confirmed by imaging)<br>• ISCD criteria for paediatric osteoporosis (or equivalent criteria), since these criteria include vertebral compression | No | Uncertain - probably varies for fracture<br>Surrogate measures are likely to be considered of limited importance | Sex hormones promote a period of rapid bone mineral accrual during puberty. Suppressing sex hormone production may slow the rate of bone mineral accrual, with potential to lead to lower bone mineral density (BMD) relative to peers. Implications for people-important outcomes (i.e. short- and longer-term fracture risk and functional impacts) need to be understood (17).<br>**Timeframe for follow-up:** during treatment (short term). After cessation of |

[10] The GRADE book and Cochrane Handbook recommend against inclusion of interim or surrogate outcomes (including radiological results such as loss of bone mineral content) which can be 'misleading since thy may not predict clinically important outcomes accurately'. Surrogates are sometimes used absence of direct evidence when there is a 'strong and well-established correlation with meaningful, patient-important outcomes'. However, for the reasons given, BMD was not prioritised.

[11] Annual monitoring of BMD is recommended during puberty suppression, as is advice to optimise bone health through lifestyle factors (physical activity, diet). While low BMD signals a need for comprehensive bone health assessment, it is not diagnostic for paediatric osteoporosis. (See O'Connell 2022; Telfer 2018)

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License.

| | Outcome domain | Specification of domain (definition, inclusions) | Example measurement methods | Primary goal of treatment | Value to those affected | Rationale |
|---|---|---|---|---|---|---|
| | | In adults, aBMD is strongly associated with fracture risk and plays a central role in diagnosing osteoporosis. This is not the case in adolescents. Interpretation of DXA-acquired measures is complicated by growth and skeletal maturation, and commonly used aged-based Z-scores (calculated relative to healthy peers) do not account for variation in pubertal timing (141). How differences in BMD during this period of rapid change in bone accrual relate to fracture risk in the short term or peak bone mass are uncertain. (8) | (crush) fractures or a clinically significant fracture history[12](110)<br>• Measures of functional impact, such as on age-appropriate normal daily activity or mobility<br>Where data are available, we will quantify fractures (e.g. fracture rate), and report fracture location and severity (e.g. Genant grading system for vertebral fractures (142)) | | | treatment, including longer term (in early and later adulthood) if effects of GnRHa can be isolated from other interventions. Factors strongly associated with bone health must be controlled for (e.g. physical activity, calcium rich diet, and risk factors).<br>**Exclusion of BMD (as a surrogate for fracture).** In studies examining intervention effects, the importance of a treatment-related difference in BMD during adolescence is uncertain. There is no well-established threshold for an important between group difference in BMD in adolescents that corresponds to an important increase in fracture risk (141, 143). In addition, there is research indicating catch-up after treatments that affect bone accrual are withdrawn (140). Understanding any effects in adulthood is therefore a priority, however, will require assessment that considers subsequent exposures which can be done more informatively in the review of GAH.[13]. |

[12] International Society for Clinical Densitometry 2019 Paediatric positions: "The finding of one or more vertebral compression (crush) fractures is indicative of osteoporosis, in the absence of local disease or high energy trauma. In the absence of vertebral compression (crush) fractures, the diagnosis of osteoporosis is indicated by the presence of both a clinically significant fracture history (3 or more long bone fractures at any age up to age 19 years) and a BMD Z-score ≤ -2.0."

[13] Reviews that report on BMD longer term (in late adolescences and early adulthood) may be relevant to the interpretation of findings, especially for understanding effects on bone health in the context of other exposures (including subsequent treatments and lifestyle factors).

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License.

| | Outcome domain | Specification of domain (definition, inclusions) | Example measurement methods | Primary goal of treatment | Value to those affected | Rationale |
|---|---|---|---|---|---|---|
| 11 | Adverse events of special interest: fertility | Reproductive function and reproductive potential.<br><br>Surrogate outcomes are eligible because direct measures of fertility (e.g. live births) are expected to be uncommon among those who have recently completed puberty suppressants.<br><br>Offering fertility counselling (FC) and access to fertility preservation (FP) are a recommended part of gender affirming care (144). Process measures such as rates of FC and FP were not prioritised for this review but may be relevant to the interpretation of findings. | Fertility measures of importance to individuals: live births, pregnancies and miscarriage.<br><br>Measures of reproductive potential (i.e. surrogates for outcomes of direct importance to individuals)<br><br>• Oocyte development<br>• Spermatogenesis | No | Uncertain, probably varies importantly | Fertility is not expected to be permanently affected by puberty suppressants (17). The value trans and gender diverse young people place on fertility relative to other outcomes may vary (145). In combination, this is why fertility is included as an important but not critical outcome for decision-making.<br><br>**Timeframe for follow-up:** after discontinuation of puberty suppressants but before starting gender affirming hormones (GAH). Effects on outcomes of importance to individuals (e.g. live births), are unlikely to be able to be isolated from other interventions (GAH, fertility treatments in early adulthood) and factors. |

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License.

# Appendix 1b. Outcomes that were considered but not prioritised for the review

The table below covers outcomes that were considered in the prioritisation process (3.1.4) but not prioritised for the systematic review of puberty suppression.

| | Outcome domain | Remarks on outcome domain | Rationale |
|---|---|---|---|
| 1 | Physical, metabolic and cognitive effects | Other systematic reviews have considered effects of puberty suppression on a range of physical, metabolic and cognitive outcomes. In the prioritisation process, there was consensus that the outcomes below were of limited importance for decision-making compared to the prioritised outcomes (e.g. quality of life, social function, adverse events of special interest).<br>• Height (depending on stage of development, the temporary halt in puberty is expected to prolong the overall duration of growth, potentially altering height) (17, 19)<br>• Body composition, weight, BMI<br>• Metabolic function, glucose metabolism, insulin sensitivity, liver and kidney function (longer term)<br>• Cognitive function (the social functioning domain covers related impacts on social functioning at home, at school and in social settings; see Appendix 1a)<br>• Hot flashes, emotional lability, mood changes (generally early in treatment) | The value young people place on height, longer-term metabolic outcomes and cognitive function relative to the other outcomes for which they seek puberty suppression is unclear.<br>• Preferences for height depend on embodiment goals (i.e. precluding meaningful interpretation of effects on average).<br>• Cognitive function (overall or on specific domains such as memory, executive function) was not prioritised because it was considered less important to young people seeking care than related impacts on social functioning (at home, at school, in other social settings) and quality of life.<br>• Any serious effects on physical, metabolic or cognitive outcomes would be captured under 'Serious adverse events' (a critical outcome) |
| 2 | Reversibility | Reversibility is considered in relation to specific outcomes. For example, a direct measure of whether puberty suppression is reversible is whether sex hormones return to expected levels for endogenous puberty after treatment ceases. | Considered in relation to specific outcomes. Since puberty suppressants are used to temporarily halt puberty, rather than induce secondary sex characteristics of the desired gender, reversibility in relation to sex characteristics was not relevant to this evidence review. |
| 3 | Discontinuation | Puberty suppressants are used temporarily, and discontinuation is a routine part of care. | Discontinuation is not a prioritised outcome for this review, because puberty suppressants are used temporarily and 'discontinuation' is therefore not an indicator of a desirable or undesirable effect. |

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License.

| | Outcome domain | Remarks on outcome domain | Rationale |
|---|---|---|---|
| | | To be addressed in the contextual information in the guideline. | |
| 4 | Reidentification | “Includes people who have discontinued a trans identity, regardless of what (if any) social or legal changes or medical or surgical treatments have been accessed during transition.” (146)<br>To be addressed in the contextual information in the guideline. | The goal of puberty suppression is to allow time for a young person to make decisions about their embodiment goals and gender affirming care. Reidentification is not prioritised as an outcome for this review because it does not indicate a desirable or undesirable effect of puberty suppression. This is underpinned by the principle that “ethical research in trans populations is founded on the understanding that any gender identity outcome (trans, binary or nonbinary, or cisgender [not transgender]) is equally acceptable, healthy, and valid.” (146) |
| 5 | Regret | Subjective experiences of ‘regret’ in relation to the gender identity experience have been defined and measured in a way that conflates this concept with other concepts (146). | To be addressed in the contextual information in the guideline, not as an outcome of this review. |

© The Author(s) for the NHMRC. 2026 Creative Commons Attribution 4.0 International License.