Effects of puberty suppression medications in young people experiencing gender incongruence or dysphoria: systematic review and meta-analysis

Supplement

23 October 2026

FOR PUBLIC CONSULTATION

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

List of supplementary material

This file includes a complete description of the methods for the review and other supplementary material as below.

Grey rows indicate material contained in this file. Blue rows indicated material in separate files.

Supplement 1. Supplementary methods S1.A: Search strategies S1.B: References to studies in existing systematic reviews S1.C: Data preparation for meta-analysis or reporting of individual effects S1.D: Differences between protocol and the review S1.E: Response to methodological review of the protocol S1.F: Response to methodological review of the systematic review report
Supplement 2. Characteristics of studies (excel file) S2A. Characteristics of included studies [excel worksheet 1] S2B. Studies assessed (excluded, awaiting assessment, ongoing and included with reasons, sources and mapping of decisions to other major systematic reviews) [excel worksheet 2]
Supplement 3. Risk of bias assessments
Supplement 4. Supplementary results
Supplement 5. Abbreviations

FOR PUBL

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

Table of Contents

List of supplementary material **2**

Supplement 1. Supplementary methods **4**

Supplement 1.A. Search strategies *4*
Ovid MEDLINE(R) ALL 1946 to March 03, 2026 4
Revised Ovid MEDLINE search 5
Ovid Embase 1974 to 2026 March 02 5
Ovid APA PsycInfo 1987 to February 2026 Week 4 6
CINAHL Ultimate (EBSCO) 6
Europe PMC 7
Trial registers 7

Supplement 1.B. References to studies in existing systematic reviews *8*

Supplement 1.C. Data preparation for meta-analysis or reporting of individual effects *15*
Effect measures 15
Data preparation by study and outcome 17

Supplement 1.D. Differences between the protocol and the review *21*

Supplement 1.E. Response to the Methodological review and other feedback *25*
Response to feedback on the protocol 25
Response to feedback on the systematic review report 28

Supplement 2. Characteristics of studies (separate Excel file) **34**

S2A. Characteristics of included studies [excel worksheet 1] 34
S2B. Studies assessed (including exclusions with reason) [excel worksheet 2] 34

Supplement 3. Risk of bias assessments **35**

Becker-Hebly 2021 35
Carmichael 2021 37
Costa 2015 38
de Vries 2011 40
Fisher 2024 41
Karamanis 2023 43
Lahaije 2025 45
Lavender 2023 46
McGregor 2024 47
Nunes-Moreno 2025 49
Schagen 2016 51
van de Grift 2020 53
van der Miesen 2020 55
Turban 2020 58
Zaitoon 2025 60

Supplement 4. Supplementary results **61**

Supplement 5. Abbreviations **62**

References used in the supplement **64**

FOR PUBLIC CONSULTATION

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

Supplement 1. Supplementary methods

Supplement 1.A. Search strategies

All database searches were run on 6 March 2026.

Ovid MEDLINE(R) ALL 1946 to March 03, 2026

#	Concept	Search Statement	Results
1	Population	Gender Dysphoria/ or Transsexualism/ or exp "Sexual and Gender Minorities"/ or Health Services for Transgender Persons/ or ((gender adj5 (dysphor* or euphor* or incongruen* or disorder* or nonconform* or non-conform* or minorit* or divers*)) or nonbinary or non-binary or trans or transgender* or transex* or transsex* or transgirl* or transboy* or transfemale* or transmale* or transfeminin* or transmasculin* or transwom* or transman or transmen or transperson* or transpeople or agender or queer or gender affirm* or gender assign* or gender reassign* or sex change or sex assign* or sex reassign* or assign* male or assign* female or birth assigned or assigned at birth or gender change or detransition* or de-transition* or reidentif* or re-identif*).ti,ab,kw.	206005
2	Intervention	exp Gonadotropin-Releasing Hormone/ or Puberty Suppression/ or Puberty Inhibitors/ or Amenorrhea/ or ((gonadotropin adj3 releasing) or gonadorelin or GnRH* or buserelin or histrelin or leuprorelin or leuprolide or Lupron or Eligard or nafarelin or triptorelin or Trelstar or goserelin or Zoladex or ((pubert* or gonadal or gonadotropin*) adj3 (suppress* or supress* or block* or halt* or suspen* or delay* or inhibit* or prevent* or stop*)) or hormon* suppression therapy or (menstrual adj (suppression or management or manipulation))).ti,ab,kw.	65769
3		1 and 2	857
4	Date	limit 3 to yr="2024 -Current"	203

Following independent peer review of the MEDLINE search strategy using the PRESS statement, the search was revised, as described in the table below. Incorporating these revisions retrieved an additional 21 records (search run on 20 March 2026). These records were imported to Covidence for screening. However, as none of these additional records was relevant, we opted not to incorporate these revisions in the other database searches.

Population terms added	transchild* or transminor* or transjuvenil* or transyoung* or transyouth* or transteen* or transtween* or transadoles* (gender* adj5 (expansive or fluid* or confus* or creative or distress* or disturb* or questioning or problem* or transition* or retransition* or re-transition* or variant* or variance)) or (dysphori* and (gender or sex*)) or (discordance and gender and sex*) or TGD or TGE or TGGE or TGN or TGNB or TGNC or gender identity or crossgender* or cross gender* or crosssex* or cross sex* or twospirit* or two-spirit* or third gender desistance

FOR PUBLIC CONSULTATION

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

Intervention terms added	postpon* or paus* luteini#ing hormone releasing pubarch* or menstrua* or menarch* or gonadarch* or spermarch* or semenarch* truncation / wildcards added: gonadotrop?in, buserelin*, hist?relin
Date	Replaced yr="2024 -Current" with: (2024* or 2025* or 2026*).yr,da,dp,dt,ed,ep,ez.
Field labels	Replaced .ti,ab,kw. with .mp. for sets 1 and 2

Revised Ovid MEDLINE search

Population	Gender Dysphoria/ or Transsexualism/ or exp "Sexual and Gender Minorities"/ or Health Services for Transgender Persons/ or ((gender adj5 (dysphor* or euphor* or incongruen* or disorder* or nonconform* or non-conform* or minorit* or divers* or expansive or fluid* or confus* or creative or distress* or disturb* or questioning or problem* or transition* or retransition* or re-transition* or variant* or variance)) or (dysphori* and (gender or sex*)) or (discordance and gender and sex*) or TGD or TGE or TGGE or TGN or TGNB or TGNC or gender identity or crossgender* or cross gender* or crosssex* or cross sex* or twospirit* or two-spirit* or third gender or nonbinary or non-binary or trans or transgender* or transex* or transsex* or transgirl* or transboy* or transfemale* or transmale* or transfeminin* or transmasculin* or transwom* or transman or transmen or transperson* or transpeople or transchild* or transminor* or transjuvenil* or transyoung* or transyouth* or transteen* or transtween* or transadoles* or agender or queer or gender affirm* or gender assign* or gender reassign* or sex change or sex assign* or sex reassign* or assign* male or assign* female or birth assigned or assigned at birth or gender change or detransition* or de-transition* or reidentif* or re-identif* or desistance).mp.
Intervention	exp Gonadotropin-Releasing Hormone/ or Puberty Suppression/ or Puberty Inhibitors/ or Amenorrhea/ or ((gonadotrop?in adj3 releasing) or gonadorelin or GnRH* or buserelin* or hist?relin or leuprorelin* or leuprolide or Lupron or Eligard or nafarelin or triptorelin or Trelstar or goserelin or Zoladex or luteini#ing hormone releasing or ((pubert* or gonadal or gonadotrop?in* or pubarch* or menstrua* or menarch* or gonadarch* or spermarch* or semenarch*) adj3 (suppress* or supress* or block* or halt* or suspen* or delay* or inhibit* or prevent* or stop* or defer* or postpon*)) or hormon* suppression therapy or (menstrual adj (suppression or management or manipulation))).mp.
Date	(2024* or 2025* or 2026*).yr,da,dp,dt,ed,ep,ez.

Ovid Embase 1974 to 2026 March 02

#	Concept	Search Statement	Results
1	Population	exp Gender Dysphoria/ or Transgenderism/ or Transgender/ or "Sexual and Gender Minority"/ or ((gender adj5 (dysphor* or euphor* or incongruen* or disorder* or nonconform* or non-conform* or minorit* or divers*)) or nonbinary or non-binary or trans or transgender* or transex* or transsex* or transgirl* or transboy* or transfemale* or transmale* or transfeminin* or transmasculin* or transwom* or transman or transmen or transperson* or transpeople or agender or queer or gender affirm* or gender assign* or gender reassign* or sex change or sex assign* or sex reassign* or assign* male or assign* female or birth assigned or assigned at birth or gender change or detransition* or de-transition* or reidentif* or re-identif*).ti,ab,kw.	263077

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

2	Intervention	Gonadorelin/ or exp Gonadorelin Agonist/ or Puberty Inhibitor/ or Puberty Suppression/ or ((gonadotropin adj3 releasing) or gonadorelin or GnRH* or buserelin or histrelin or leuprorelin or leuprolide or Lupron or Eligard or nafarelin or triptorelin or Trelstar or goserelin or Zoladex or ((pubert* or gonadal or gonadotropin*) adj3 (suppress* or supress* or block* or halt* or suspen* or delay* or inhibit* or prevent* or stop*)) or hormon* suppression therapy or (menstrual adj (suppression or management or manipulation))).ti,ab,kw.	97828
3		1 and 2	1784
4	Date	limit 3 to yr="2024 -Current"	411

Ovid APA PsycInfo 1987 to February 2026 Week 4

#	Concept	Search Statement	Results
1	Population	Gender Dysphoria/ or Gender Nonbinary/ or Gender Nonconforming/ or Transgender/ or Transsexualism/ or Gender Transition/ or ((gender adj5 (dysphor* or euphor* or incongruen* or disorder* or nonconform* or non-conform* or minorit* or divers*)) or nonbinary or non-binary or trans or transgender* or transex* or transsex* or transgirl* or transboy* or transfemale* or transmale* or transfeminin* or transmasculin* or transwom* or transman or transmen or transperson* or transpeople or agender or queer or gender affirm* or gender assign* or gender reassign* or sex change or sex assign* or sex reassign* or assign* male or assign* female or birth assigned or assigned at birth or gender change or detransition* or de-transition* or reidentif* or re-identif*).mp.	49325
2	Intervention	exp Gonadotropic Hormones/ or Amenorrhea/ or ((gonadotropin adj3 releasing) or gonadorelin or GnRH* or buserelin or histrelin or leuprorelin or leuprolide or Lupron or Eligard or nafarelin or triptorelin or Trelstar or goserelin or Zoladex or ((pubert* or gonadal or gonadotropin*) adj3 (suppress* or supress* or block* or halt* or suspen* or delay* or inhibit* or prevent* or stop*)) or hormon* suppression therapy or (menstrual adj (suppression or management or manipulation))).mp.	5379
3		1 and 2	320
4	Date	limit 3 to yr="2024 -Current"	69

CINAHL Ultimate (EBSCO)

6 March 2026

#	Search Statement	Results
1	(MH "Gender Transition" OR MH "Gender Dysphoria" OR MH "Gender Identity+" OR MH "Gender-Nonconforming Persons+" OR MH "Sexual and Gender Minorities+" OR MH "Gender Affirmation Procedures+" OR TX ((gender N5 (dysphor* or euphor* or incongruen* or disorder* or nonconform* or non-conform* or minorit* or divers*)) OR nonbinary or non-binary or trans or transgender* or transex* or transsex* or transgirl* or transboy* or transfemale* or transmale* or transfeminin* or transmasculin* or transwom* or transman or transmen or transperson* or transpeople or agender or queer or gender affirm* or gender assign* or gender reassign* or sex change or sex assign* or sex reassign* or assign* male or assign* female or birth assigned or assigned at birth or gender change or detransition* or de-transition* or reidentif* or re-identif*)) AND (MH "Gonadotropin-Releasing Hormone+" OR MH "Puberty	431

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

	Suppression" OR MH "Amenorrhea" OR ((gonadotropin N3 releasing) or gonadorelin or GnRH* or buserelin or histrelin or leuprorelin or leuprolide or Lupron or Eligard or nafarelin or triptorelin or Trelstar or goserelin or Zoladex or (pubert* or gonadal or gonadotropin*) N3 (suppress* or supress* or block* or halt* or suspen* or delay* or inhibit* or prevent* or stop*)) or hormon* suppression therapy or (menstrual adj (suppression or management or manipulation)))	
2	Limit to publication date – 2026 01/01/2024 - 03/31/2026	99

Europe PMC

6 March 2026
TITLE_ABS:("puberty blockers" OR "puberty suppression" OR "puberty supression" OR "puberty inhibitors" OR "pubertal blockers" OR "pubertal suppression" OR "pubertal supression" OR "pubertal inhibitors" OR "halt puberty" OR "suspend puberty" OR "delay puberty" OR "inhibit puberty" OR "delay puberty" OR "prevent puberty" OR "stop puberty" OR (("gonadotropin releasing" OR GnRH*) AND pubert*) OR "menstrual suppression" OR "menstrual management" OR "menstrual manipulation") AND SRC:(PPR) NOT TITLE:(mice OR mouse OR rats OR precocious)
Limited to preprints from 2024 – 2026
Records retrieved = 27

Trial registers

Searched 31 March 2026. No date limit was applied.

ClinicalTrials.gov (https://clinicaltrials.gov/)

Condition/disease: dysphoria OR dysphoric OR incongruence OR incongruent OR gender identity OR transgender
Intervention/treatment: puberty OR pubertal OR gonadotropin OR gonadotrophin OR GnRH OR GnRHa OR gonadal
Results: 41 records

WHO International Clinical Trials Registry Platform (https://trialsearch.who.int/)

Condition: dysphoria OR dysphoric OR incongruence OR incongruent OR gender identity OR transgender
Intervention: puberty OR pubertal OR gonadotropin OR gonadotrophin OR GnRH OR GnRHa OR gonadal
Results: 1 record

ISRCTN (https://www.isrctn.com/)

(dysphoria OR dysphoric OR incongruence OR incongruent OR gender identity OR transgender) AND (puberty OR pubertal OR gonadotropin OR gonadotrophin OR GnRH OR GnRHa OR gonadal)
Results: 3 records (PATHWAYS trial ISRCTN12491684 (main trial) and ISRCTN69296987 (longitudinal observational study); ISRCTN98564473 Dr Eilis Kennedy – no GnRHa intervention)
Results: 3 records

FOR PUBLIC CONSULTATION

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

Supplement 1.B. References to studies in existing systematic reviews

See Supplement 2 (excel spreadsheet) for a filterable list of all studies, eligibility for this and other systematic reviews.

Achille C, Taggart T, Eaton NR, Osipoff J, Tafuri K, Lane A, et al. Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: preliminary results. Int J Pediatr Endocrinol. 2020;2020:8. http://dx.doi.org/10.1186/s13633-020-00078-2

Akgul S, Bonny AE, Ford N, Holland-Hall C, Chelvakumar G. Experiences of gender minority youth with the intrauterine system. J Adolesc Health. 2019;65(1):32–8. http://dx.doi.org/10.1016/j.jadohealth.2018.11.010

Allen LR, Watson LB, Egan AM, Moser CN. Well-being and suicidality among transgender youth after gender-affirming hormones. Clinical Practice in Pediatric Psychology. 2019;7(3):302–11. http://dx.doi.org/10.1037/cpp0000288

Amir H, Perl L, Barda S, Lantsberg D, Sege Becker A, Israeli G, et al. Semen quality and cryopreservation in adolescent transgender females. Hum Reprod. 2021;36(SUPPL 1):i336–i7. http://dx.doi.org/10.1093/humrep/deab130.440

Arcelus J, Claes L, Witcomb GL, Marshall E, Bouman WP. Risk factors for non-suicidal self-injury among trans youth. J Sex Med. 2016;13(3):402–12. http://dx.doi.org/10.1016/j.jsxm.2016.01.003

Arnoldussen M, Hooijman EC, Kreukels BP, de Vries AL. Association between pre-treatment IQ and educational achievement after gender-affirming treatment including puberty suppression in transgender adolescents. Clin Child Psychol Psychiatry. 2022;27(4):1069–76. http://dx.doi.org/10.1177/13591045221091652

Becker I, Auer M, Barkmann C, Fuss J, Möller B, Nieder TO, et al. A cross-sectional multicenter study of multidimensional body image in adolescents and adults with gender dysphoria before and after transition-related medical interventions. Arch Sex Behav. 2018;47(8):2335–47. http://dx.doi.org/10.1007/s10508-018-1278-4

Boogers LS, Wiepjes CM, Klink DT, Hellinga I, van Trotsenburg ASP, den Heijer M, et al. Transgender girls grow tall: adult height is unaffected by GnRH analogue and estradiol treatment. J Clin Endocrinol Metab. 2022;107(9):e3805–e15. http://dx.doi.org/10.1210/clinem/dgac349

Boogers LS, van der Loos M, Wiepjes CM, van Trotsenburg ASP, den Heijer M, Hannema SE. The dose-dependent effect of estrogen on bone mineral density in trans girls. Eur J Endocrinol. 2023;189(2):290–6. http://dx.doi.org/10.1093/ejendo/lvad116

Boogers LS, Reijtenbagh SJP, Wiepjes CM, van Trotsenburg ASP, den Heijer M, Hannema SE. Time course of body composition changes in transgender adolescents during puberty suppression and sex hormone treatment. J Clin Endocrinol Metab. 2024;109(8):e1593–e601. http://dx.doi.org/10.1210/clinem/dgad750

Boogers LS, Sardo Infirri SA, Bouchareb A, Dijkman BAM, Helder D, de Blok CJM, et al. Variations in volume: breast size in trans women in relation to timing of testosterone suppression. J Clin Endocrinol Metab. 2025;110(5):e1404–e10. http://dx.doi.org/10.1210/clinem/dgae573

Burke SM, van Heesewijk JO, Menks WM, Klink DT, Kreukels BPC, Cohen-Kettenis PT, et al. Postnatal effects of sex hormones on click-evoked otoacoustic emissions: a study of adolescents with gender dysphoria. Arch Sex Behav. 2020;49(2):455–65. http://dx.doi.org/10.1007/s10508-020-01652-8

Cantu AL, Moyer DN, Connelly KJ, Holley AL. Changes in anxiety and depression from intake to first follow-up among transgender youth in a pediatric endocrinology clinic. Transgend Health. 2020;5(3):196–200. http://dx.doi.org/10.1089/trgh.2019.0077

FOR PUBLIC CONSULTATION

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

Chelliah P, Lau M, Kuper LE. Changes in gender dysphoria, interpersonal minority stress, and mental health among transgender youth after one year of hormone therapy. J Adolesc Health. 2024;74(6):1106–11. http://dx.doi.org/10.1016/j.jadohealth.2023.12.024

Chen M, Fuqua J, Eugster EA. Characteristics of referrals for gender dysphoria over a 13-year period. J Adolesc Health. 2016;58(3):369–71. http://dx.doi.org/10.1016/j.jadohealth.2015.11.010

Chen D, Abrams M, Clark L, Ehrensaft D, Tishelman AC, Chan YM, et al. Psychosocial characteristics of transgender youth seeking gender-affirming medical treatment: baseline findings from the Trans Youth Care Study. J Adolesc Health. 2021;68:1104–11. https://dx.doi.org/10.1016/j.jadohealth.2020.07.033

Chen D, Berona J, Chan YM, Ehrensaft D, Garofalo R, Hidalgo MA, et al. Psychosocial functioning in transgender youth after 2 years of hormones. N Engl J Med. 2023;388(3):240–50. http://dx.doi.org/10.1056/NEJMoa2206297

Chiniara LN, Bonifacio HJ, Palmert MR. Characteristics of adolescents referred to a gender clinic: are youth seen now different from those in initial reports? Horm Res Paediatr. 2018;89(6):434–41. http://dx.doi.org/10.1159/000489608

Ciancia S, Klink D, Craen M, Cools M. Early puberty suppression and gender-affirming hormones do not alter final height in transgender adolescents. Eur J Endocrinol. 2023;189(3):396–401. http://dx.doi.org/10.1093/ejendo/lvad125

Cohen-Kettenis PT, Schagen SE, Steensma TD, de Vries AL, Delemarre-van de Waal HA. Puberty suppression in a gender-dysphoric adolescent: a 22-year follow-up. Arch Sex Behav. 2011;40(4):843–7. http://dx.doi.org/10.1007/s10508-011-9758-9

Costa R, Carmichael P, Colizzi M. To treat or not to treat: puberty suppression in childhood-onset gender dysphoria. Nat Rev Urol. 2016;13(8):456–62. http://dx.doi.org/10.1038/nrurol.2016.128

de Nie I, Mulder CL, Meißner A, Schut Y, Holleman EM, van der Sluis WB, et al. Histological study on the influence of puberty suppression and hormonal treatment on developing germ cells in transgender women. Hum Reprod. 2022;37(2):297–308. http://dx.doi.org/10.1093/humrep/deab240

Delemarre-Van De Waal HA, Cohen-Kettenis PT. Clinical management of gender identity disorder in adolescents: a protocol on psychological and paediatric endocrinology aspects. European Journal of Endocrinology, Supplement. 2006;155(1):S131–S7. http://dx.doi.org/https://dx.doi.org/10.1530/eje.1.02231

Dittrich R, Binder H, Cupisti S, Hoffmann I, Beckmann MW, Mueller A. Endocrine treatment of male-to-female transsexuals using gonadotropin-releasing hormone agonist. Exp Clin Endocrinol Diabetes. 2005;113(10):586–92. http://dx.doi.org/10.1055/s-2005-865900

Donaldson AA, Hall A, Neukirch J, Kasper V, Simones S, Gagnon S, et al. Multidisciplinary care considerations for gender nonconforming adolescents with eating disorders: a case series. Int J Eat Disord. 2018;51(5):475–9. http://dx.doi.org/10.1002/eat.22868

Eitel KB, Hodax JK, DiVall S, Kidd KM, Salehi P, Sequeira GM. Leuprolide acetate for puberty suppression in transgender and gender diverse youth: a comparison of subcutaneous Eligard versus intramuscular Lupron. J Adolesc Health. 2023;72(2):307–11. http://dx.doi.org/10.1016/j.jadohealth.2022.09.017

Elkadi J, Chudleigh C, Maguire AM, Ambler GR, Scher S, Kozlowska K. Developmental pathway choices of young people presenting to a gender service with gender distress: a prospective follow-up study. Children (Basel). 2023;10(2). http://dx.doi.org/10.3390/children10020314

FOR PUBLIC CONSULTATION

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

Fontanari AMV, Vilanova F, Schneider MA, Chinazzo I, Soll BM, Schwarz K, et al. Gender affirmation is associated with transgender and gender nonbinary youth mental health improvement. LGBT Health. 2020;7(5):237–47. http://dx.doi.org/10.1089/lgbt.2019.0046

Ghelani R, Lim C, Brain C, Fewtrell M, Butler G. Sudden sex hormone withdrawal and the effects on body composition in late pubertal adolescents with gender dysphoria. J Pediatr Endocrinol Metab. 2020;33(1):107–12. http://dx.doi.org/10.1515/jpem-2019-0045

Grimstad FW, Knoll MM, Jacobson JD. Oxandrolone use in trans-masculine youth appears to increase adult height: preliminary evidence. LGBT Health. 2021;8(4):300–6. http://dx.doi.org/10.1089/lgbt.2020.0355

Hisle-Gorman E, Schvey NA, Adirim TA, Rayne AK, Susi A, Roberts TA, et al. Mental healthcare utilization of transgender youth before and after affirming treatment. J Sex Med. 2021;18(8):1444–54. http://dx.doi.org/10.1016/j.jsxm.2021.05.014

Hobson BJ, Lett E, Hawkins LA, Swendiman RA, Nance ML, Dowshen NL. Transgender youth experiences with implantable GnRH agonists for puberty suppression. Transgend Health. 2022;7(4):364–8. http://dx.doi.org/10.1089/trgh.2021.0006

Jarin J, Pine-Twaddell E, Trotman G, Stevens J, Conard LA, Tefera E, et al. Cross-sex hormones and metabolic parameters in adolescents with gender dysphoria. Pediatrics. 2017;139(5). http://dx.doi.org/10.1542/peds.2016-3173

Jensen RK, Jensen JK, Simons LK, Chen D, Rosoklija I, Finlayson CA. Effect of concurrent gonadotropin-releasing hormone agonist treatment on dose and side effects of gender-affirming hormone therapy in adolescent transgender patients. Transgend Health. 2019;4(1):300–3. http://dx.doi.org/10.1089/trgh.2018.0061

Joseph T, Ting J, Butler G. The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort. J Pediatr Endocrinol Metab. 2019;32(10):1077–81. http://dx.doi.org/10.1515/jpem-2019-0046

Karakılıç Özturan E, Öztürk AP, Baş F, Erdoğdu AB, Kaptan S, Kardelen Al AD, et al. Endocrinological approach to adolescents with gender dysphoria: experience of a pediatric endocrinology department in a tertiary center in Turkey. J Clin Res Pediatr Endocrinol. 2023;15(3):276–84. http://dx.doi.org/10.4274/jcrpe.galenos.2023.2023-1-13

Khatchadourian K, Amed S, Metzger DL. Clinical management of youth with gender dysphoria in Vancouver. J Pediatr. 2014;164(4):906–11. http://dx.doi.org/10.1016/j.jpeds.2013.10.068

Klaver M, de Mutsert R, Wiepjes CM, Twisk JWR, den Heijer M, Rotteveel J, et al. Early hormonal treatment affects body composition and body shape in young transgender adolescents. J Sex Med. 2018;15(2):251–60. http://dx.doi.org/10.1016/j.jsxm.2017.12.009

Klaver M, de Mutsert R, van der Loos M, Wiepjes CM, Twisk JWR, den Heijer M, et al. Hormonal treatment and cardiovascular risk profile in transgender adolescents. Pediatrics. 2020;145(3). http://dx.doi.org/10.1542/peds.2019-0741

Klink D, Caris M, Heijboer A, van Trotsenburg M, Rotteveel J. Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. J Clin Endocrinol Metab. 2015;100(2):E270–5. http://dx.doi.org/10.1210/jc.2014-2439

Kuper LE, Stewart S, Preston S, Lau M, Lopez X. Body Dissatisfaction and mental health outcomes of youth on gender-affirming hormone therapy. Pediatrics. 2020;145(4). http://dx.doi.org/10.1542/peds.2019-3006

FOR PUBLIC CONSULTATION

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

Lee JY, Finlayson C, Olson-Kennedy J, Garofalo R, Chan YM, Glidden DV, et al. Low bone mineral density in early pubertal transgender/gender diverse youth: findings from the Trans Youth Care Study. J Endocr Soc. 2020;4(9):bvaa065. http://dx.doi.org/10.1210/jendso/bvaa065

Lopez CM, Solomon D, Boulware SD, Christison-Lagay ER. Trends in the use of puberty blockers among transgender children in the United States. J Pediatr Endocrinol Metab. 2018;31(6):665–70. http://dx.doi.org/10.1515/jpem-2018-0048

López de Lara D, Pérez Rodríguez O, Cuellar Flores I, Pedreira Masa JL, Campos-Muñoz L, Cuesta Hernández M, et al. [Psychosocial assessment in transgender adolescents]. An Pediatr (Engl Ed). 2020;93(1):41–8. http://dx.doi.org/10.1016/j.anpedi.2020.01.019

Lynch MM, Khandheria MM, Meyer Iii WJ. Retrospective study of the management of childhood and adolescent gender identity disorder using medroxyprogesterone acetate. International Journal of Transgenderism. 2015;16(4):201–8. http://dx.doi.org/10.1080/15532739.2015.1080649

McPherson S, Freedman DE. Psychological outcomes of 12-15 year olds with gender dysphoria receiving pubertal suppression: assessing reliable change and recovery. medRxiv. 2023:2023.05. 30.23290763. http://dx.doi.org/10.1101/2023.05.30.23290763

Mejia-Otero JD, White P, Lopez X. Effectiveness of puberty suppression with gonadotropin-releasing hormone agonists in transgender youth. Transgend Health. 2021;6(1):31–5. http://dx.doi.org/10.1089/trgh.2020.0007

Moussaoui D, O'Connell MA, Elder CV, Grover SR, Pang KC. Characteristics of menstrual suppression and its association with mental health in transgender adolescents. Obstet Gynecol. 2023;142(5):1096–104. http://dx.doi.org/10.1097/aog.0000000000005317

Mullins ES, Geer R, Metcalf M, Piccola J, Lane A, Conard LAE, et al. Thrombosis risk in transgender adolescents receiving gender-affirming hormone therapy. Pediatrics. 2021;147(4). http://dx.doi.org/10.1542/peds.2020-023549

Muns-Aponte S, Dell AW, Geist C, Mihalopoulos N. Mental health profiles among transgender adolescents & young adults. J Adolesc Health. 2019;64(2 Supplement):S99. http://dx.doi.org/10.1016/j.jadohealth.2018.10.211

Nahata L, Quinn GP, Caltabellotta NM, Tishelman AC. Mental health concerns and insurance denials among transgender adolescents. LGBT Health. 2017;4(3):188–93. http://dx.doi.org/10.1089/lgbt.2016.0151

Nasomyont N, Meisman AR, Ecklund K, Vajapeyam S, Cecil KM, Tkach JA, et al. Changes in bone marrow adipose tissue in transgender and gender non-conforming youth undergoing pubertal suppression: a pilot study. J Clin Densitom. 2022;25(4):485–9. http://dx.doi.org/10.1016/j.jocd.2022.06.006

Navabi B, Tang K, Khatchadourian K, Lawson ML. Pubertal suppression, bone mass, and body composition in youth with gender dysphoria. Pediatrics. 2021;148(4). http://dx.doi.org/10.1542/peds.2020-039339

Neyman A, Fuqua JS, Eugster EA. Bicalutamide as an androgen blocker with secondary effect of promoting feminization in male-to-female transgender adolescents. J Adolesc Health. 2019;64(4):544–6. http://dx.doi.org/10.1016/j.jadohealth.2018.10.296

Ni J, Chi C, Aye T. Review of implant gonadotrophin-releasing hormone agonist use: experience in a single academic center. Horm Res Paediatr. 2023;96(5):523–6. http://dx.doi.org/10.1159/000529733

Nokoff NJ, Scarbro SL, Moreau KL, Zeitler P, Nadeau KJ, Juarez-Colunga E, et al. Body composition and markers of cardiometabolic health in transgender youth compared with cisgender youth. J Clin Endocrinol Metab. 2020;105(3):e704–14. http://dx.doi.org/10.1210/clinem/dgz029

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

FOR PUBLIC CONSULTATION

Nokoff NJ, Scarbro SL, Moreau KL, Zeitler P, Nadeau KJ, Reirden D, et al. Body composition and markers of cardiometabolic health in transgender youth on gonadotropin-releasing hormone agonists. Transgend Health. 2021;6(2):111–9. http://dx.doi.org/10.1089/trgh.2020.0029

Olsavsky AL, Grannis C, Bricker J, Chelvakumar G, Indyk JA, Leibowitz SF, et al. Associations among gender-affirming hormonal interventions, social support, and transgender adolescents' mental health. J Adolesc Health. 2023;72(6):860–8. http://dx.doi.org/10.1016/j.jadohealth.2023.01.031

Olson-Kennedy J, Streeter LH, Garofalo R, Chan YM, Rosenthal SM. Histrelin implants for suppression of puberty in youth with gender dysphoria: a comparison of 50 mcg/day (Vantas) and 65 mcg/day (SupprelinLA). Transgend Health. 2021;6(1):36–42. http://dx.doi.org/10.1089/trgh.2020.0055

Pang KC, Peri AJS, Chung HE, Telfer M, Elder CV, Grover S, et al. Rates of fertility preservation use among transgender adolescents. JAMA Pediatr. 2020;174(9):890–1. http://dx.doi.org/10.1001/jamapediatrics.2020.0264

Peri A, Ahler A, Gook D, O'Connell MA, Bourne H, Nightingale M, et al. Predicting successful sperm retrieval in transfeminine adolescents after testicular biopsy. J Assist Reprod Genet. 2021;38(10):2735–43. http://dx.doi.org/10.1007/s10815-021-02293-z

Perl L, Segev-Becker A, Israeli G, Elkon-Tamir E, Oren A. Blood pressure dynamics after pubertal suppression with gonadotropin-releasing hormone analogs followed by testosterone treatment in transgender male adolescents: a pilot study. LGBT Health. 2020;7(6):340–4. http://dx.doi.org/10.1089/lgbt.2020.0026

Perl L, Elkon-Tamir E, Segev-Becker A, Israeli G, Brener A, Oren A. Blood pressure dynamics after pubertal suppression with gonadotropin-releasing hormone analogs followed by estradiol treatment in transgender female adolescents: a pilot study. J Pediatr Endocrinol Metab. 2021;34(6):741–5. http://dx.doi.org/10.1515/jpem-2021-0172

Pine-Twaddell E, Newfield RS, Marinkovic M. Extended use of histrelin implant in pediatric patients. Transgend Health. 2023;8(3):264–72. http://dx.doi.org/10.1089/trgh.2021.0130

Pullen Sansfaçon A, Temple-Newhook J, Suerich-Gulick F, Feder S, Lawson ML, Ducharme J, et al. The experiences of gender diverse and trans children and youth considering and initiating medical interventions in Canadian gender-affirming speciality clinics. Int J Transgend. 2019;20(4):371–87. http://dx.doi.org/10.1080/15532739.2019.1652129

Roberge S, Roberge T, Corathers S, Nasomyont N. Determinants of bone mass accrual in transgender and gender diverse youth undergoing pubertal suppression therapy. J Clin Densitom. 2024;27(3):101505. http://dx.doi.org/10.1016/j.jocd.2024.101505

Roy MK, Bothwell S, Kelsey MM, Ma NS, Moreau KL, Nadeau KJ, et al. Bone density in transgender youth on gender-affirming hormone therapy. J Endocr Soc. 2024;8(5):bvae045. http://dx.doi.org/10.1210/jendso/bvae045

Russell I, Pearson B, Masic U. A Longitudinal study of features associated with autism spectrum in clinic referred, gender diverse adolescents accessing puberty suppression treatment. J Autism Dev Disord. 2021;51(6):2068–76. http://dx.doi.org/10.1007/s10803-020-04698-8

Schagen SEE, Lustenhouwer P, Cohen-Kettenis PT, Delemarre-van de Waal HA, Hannema SE. Changes in adrenal androgens during puberty suppression and gender-affirming hormone treatment in adolescents with gender dysphoria. J Sex Med. 2018;15(9):1357–63. http://dx.doi.org/10.1016/j.jsxm.2018.07.017

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. http://dx.doi.org/10.1210/clinem/dgaa604

FOR PUBLIC CONSULTATION

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

Schneider MA, Spritzer PM, Soll BMB, Fontanari AMV, Carneiro M, Tovar-Moll F, et al. Brain maturation, cognition and voice pattern in a gender dysphoria case under pubertal suppression. Front Hum Neurosci. 2017;11:528. http://dx.doi.org/10.3389/fnhum.2017.00528

Schulmeister C, Millington K, Kaufman M, Finlayson C, Kennedy JO, Garofalo R, et al. Growth in transgender/gender-diverse youth in the first year of treatment with gonadotropin-releasing hormone agonists. J Adolesc Health. 2022;70(1):108–13. http://dx.doi.org/10.1016/j.jadohealth.2021.06.022

Schwartz BI, Bear B, Short VL, Kazak AE. Outcomes of Menstrual Management Use in Transgender and Gender-Diverse Adolescents. Obstet Gynecol. 2023;141(4):748–55. http://dx.doi.org/10.1097/aog.0000000000005123

Segev-Becker A, Israeli G, Elkon-Tamir E, Perl L, Sekler O, Amir H, et al. Children and adolescents with gender dysphoria in Israel: increasing referral and fertility preservation rates. Endocr Pract. 2020;26(4):423–8. http://dx.doi.org/10.4158/ep-2019-0418

Sequeira GM, Kidd K, El Nokali NE, Rothenberger SD, Levine MD, Montano GT, et al. Early effects of testosterone initiation on body mass index in transmasculine adolescents. J Adolesc Health. 2019;65(6):818–20. http://dx.doi.org/10.1016/j.jadohealth.2019.06.009

Staphorsius AS, Kreukels BP, Cohen-Kettenis PT, Veltman DJ, Burke SM, Schagen SE, et al. Puberty suppression and executive functioning: An fMRI-study in adolescents with gender dysphoria. Psychoneuroendocrinology. 2015;56:190–9. http://dx.doi.org/10.1016/j.psyneuen.2015.03.007

Stoffers IE, de Vries MC, Hannema SE. Physical changes, laboratory parameters, and bone mineral density during testosterone treatment in adolescents with gender dysphoria. J Sex Med. 2019;16(9):1459–68. http://dx.doi.org/10.1016/j.jsxm.2019.06.014

Strang JF, Chen D, Nelson E, Leibowitz SF, Nahata L, Anthony LG, et al. Transgender youth executive functioning: relationships with anxiety symptoms, autism spectrum disorder, and gender-affirming medical treatment status. Child Psychiatry Hum Dev. 2022;53(6):1252–65. http://dx.doi.org/10.1007/s10578-021-01195-6

Tack LJ, Craen M, Dhondt K, Vanden Bossche H, Laridaen J, Cools M. Consecutive lynestrenol and cross-sex hormone treatment in biological female adolescents with gender dysphoria: a retrospective analysis. Biol Sex Differ. 2016;7:14. http://dx.doi.org/10.1186/s13293-016-0067-9

Tack LJW, Heyse R, Craen M, Dhondt K, Bossche HV, Laridaen J, et al. Consecutive cyproterone acetate and estradiol treatment in late-pubertal transgender female adolescents. J Sex Med. 2017;14(5):747–57. http://dx.doi.org/10.1016/j.jsxm.2017.03.251

Tack LJW, Craen M, Lapauw B, Goemaere S, Toye K, Kaufman JM, et al. Proandrogenic and antiandrogenic progestins in transgender youth: differential effects on body composition and bone metabolism. J Clin Endocrinol Metab. 2018;103(6):2147–56. http://dx.doi.org/10.1210/jc.2017-02316

Tordoff DM, Wanta JW, Collin A, Stepney C, Inwards-Breland DJ, Ahrens K. Mental health outcomes in transgender and nonbinary youths receiving gender-affirming care. JAMA Netw Open. 2022;5(2):e220978. http://dx.doi.org/10.1001/jamanetworkopen.2022.0978

Valentine A, Davis S, Furniss A, Dowshen N, Kazak AE, Lewis C, et al. Multicenter analysis of cardiometabolic-related diagnoses in transgender and gender-diverse youth: A PEDSnet Study. J Clin Endocrinol Metab. 2022;107(10):e4004–e14. http://dx.doi.org/10.1210/clinem/dgac469

van der Loos MA, Hellinga I, Vlot MC, Klink DT, den Heijer M, Wiepjes CM. Development of hip bone geometry during gender-affirming hormone therapy in transgender adolescents resembles that of the experienced gender when

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

FOR PUBLIC CONSULTATION

pubertal suspension is started in early puberty. J Bone Miner Res. 2021;36(5):931–41. http://dx.doi.org/10.1002/jbmr.4262

van der Loos M, Klink DT, Hannema SE, Bruinsma S, Steensma TD, Kreukels BPC, et al. Children and adolescents in the Amsterdam Cohort of Gender Dysphoria: trends in diagnostic- and treatment trajectories during the first 20 years of the Dutch Protocol. J Sex Med. 2023;20(3):398–409. http://dx.doi.org/10.1093/jsxmed/qdac029

van der Loos M, Vlot MC, Klink DT, Hannema SE, den Heijer M, Wiepjes CM. Bone mineral density in transgender adolescents treated with puberty suppression and subsequent gender-affirming hormones. JAMA Pediatr. 2023;177(12):1332–41. http://dx.doi.org/10.1001/jamapediatrics.2023.4588

van Heesewijk J, Steenwijk MD, Kreukels BPC, Veltman DJ, Bakker J, Burke SM. Alterations in the inferior fronto-occipital fasciculus - a specific neural correlate of gender incongruence? Psychol Med. 2023;53(8):3461–70. http://dx.doi.org/10.1017/s0033291721005547

Vlot MC, Klink DT, den Heijer M, Blankenstein MA, Rotteveel J, Heijboer AC. Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone. 2017;95:11–9. http://dx.doi.org/10.1016/j.bone.2016.11.008

Waldner RC, Doulla M, Atallah J, Rathwell S, Grimbly C. Leuprolide acetate and QTc interval in gender-diverse youth. Transgend Health. 2023;8(1):84–8. http://dx.doi.org/10.1089/trgh.2021.0102

Willemsen LA, Boogers LS, Wiepjes CM, Klink DT, van Trotsenburg ASP, den Heijer M, et al. Just as tall on testosterone; a neutral to positive effect on adult height of GnRHa and testosterone in trans boys. J Clin Endocrinol Metab. 2023;108(2):414–21. http://dx.doi.org/10.1210/clinem/dgac571

Wittlin NM, Gallagher NM, Atwood S, Olson KR. Mental health during medical transition in a US and Canadian sample of early socially transitioned transgender youth. J Adolesc Health. 2025;76(2):228–37. http://dx.doi.org/10.1016/j.jadohealth.2024.10.023

FOR PUBLIC CONSULTATION

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

Supplement 1.C. Data preparation for meta-analysis or reporting of individual effects

In this document we report the details of the data preparation for meta-analysis or reporting of individual effects following data extraction, including any algebraic manipulations and imputations required to calculate the effect estimates and their standard errors. We also report details of the effect measures. Details of the meta-analysis methods are available in the report.

Abbreviations used in this section are as follows.

Abbreviation	Full term / explanation
CI	Confidence interval
CI_{upper}	Confidence interval upper limit
CI_{lower}	Confidence interval lower limit
Correlation	Correlation between the baseline and follow-up measure
J	Small-sample correction
SD	Standard deviation
SMD	Standardised mean difference
q1	25th percentile
q3	75th percentile

Effect measures

The effect estimates (and standard errors) were calculated using the function `meta esize` in Stata version 19. Here, we describe specific instances where we did not use the inbuilt function for calculating the standardised mean difference (SMD). All SMDs used the Hedges' adjusted *g* version. For proportions, we applied the Freeman–Tukey double-arcsine variance-stabilising transformation and meta-analysed the transformed estimates (1, 2).

Standardised mean difference

We calculated the standardised mean difference (SMD) for **uncontrolled before-after studies** using the following formulae:

$$SMD = J * \left(\frac{mean_{after} - mean_{before}}{\sqrt{\frac{SD^2_{after} + SD^2_{before}}{2}}} \right),$$

where $J =$ small-sample correction $= 1 - \frac{3}{4(N-2)-1}$, and

$$SE_{SMD} = \sqrt{\frac{1}{N} + \frac{SMD^2}{2*N}} \times \sqrt{2(1 - correlation)} \times J.$$

N is the number of participants. In instances when the number of participants at follow-up differed to baseline, we assumed the follow-up sample size. The correlation was assumed to be 0.5 for the primary analysis, and we investigated the impact of assuming a range of correlations (0, 0.3, 0.7) as part of the sensitivity analyses.

We calculated partial SMDs from regression models for studies with **an independent control group** using the following formulae (3):

FOR PUBLIC CONSULTATION

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

$$SMD = J\left(\frac{\hat{\beta}_{treat}}{SD_{pooled}}\right),$$

where $\hat{\beta}_1$ = estimate of intervention effect from the regression model, $J =$ small-sample correction $= 1 - \frac{3}{4(df)-1}$, with $df = N - C - 1$, N = total sample size, C = number of independent variables in the model, and,

$$SD_{pooled} = \sqrt{\frac{(n_{treat} - 1)SD_{treat}^2 + (n_{comparison} - 1)SD_{comparison}^2}{n_{treat} + n_{comparison} - 2}}.$$

$$SE_{SMD} = \sqrt{\frac{n_{treat} + n_{comparison}}{n_{treat} \times n_{comparison}} + \frac{SMD^2}{2 \times (n_{treat} + n_{comparison})}} \times J$$

FOR PUBLIC CONSULTATION

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

Data preparation by study and outcome

Becker-Hebly 2021

Quality of life

No algebraic manipulations or imputations were required.

Social functioning

The direction of the scale was reversed to match the direction of the scales from the other included studies. This was achieved by multiplying the SMD by -1.

Mental health – psychological functioning

No algebraic manipulations or imputations were required.

Carmichael 2021

Gender dysphoria

Means and standard deviations were calculated from reported percentiles using the following formulae:

Data extracted	Mean calculation	SD calculation	Reference
N q1 = 25th percentile median q3 = 75th percentile	$\frac{(q1 + median + q3)}{3}$	$SD = \frac{q_3 - q_1}{2\phi_2}$ where $\Phi_2 = invnormal\left(\frac{0.75n - 0.125}{n + 0.25}\right)$	Equations 14 and 16 from (4).

The formulae in Section *Standardised mean difference* were then applied to calculate the SMD and its SE for an uncontrolled before-after study.

Body Image

Reported data were means and confidence intervals. From the confidence intervals, standard deviations were calculated using the following formulae (from Section 6.3 of the *Cochrane Handbook for Systematic Reviews of Interventions* (5)):

$$SE = \frac{(CI_{upper} - CI_{lower})}{2(invttail(n-1, 0.025))} \text{ and } SD = SE \times \sqrt{n}$$

The formulae in Section *Standardised mean difference* were then applied to calculate the SMD and its SE for an uncontrolled before-after study.

Menstruation

The proportion of youth menstruating at follow-up who were menstruating at baseline was calculated. The Freeman-Tukey double-arcsine variance stabilising transformation was applied prior to meta-analysis.

Costa 2015

Social functioning

No algebraic manipulations or imputations were required.

FOR PUBLIC CONSULTATION

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

de Vries 2011

Gender dysphoria

The formulae in Section *Standardised mean difference* were applied to calculate the SMD and its SE for an uncontrolled before-after study.

Body image

The formulae in Section *Standardised mean difference* were applied to calculate the SMD and its SE for an uncontrolled before-after study.

Fisher 2024

Menstruation

The proportion of youth menstruating at follow-up who were menstruating at baseline was calculated. The Freeman-Tukey double-arcsine variance stabilising transformation was applied prior to meta-analysis.

Karamanis 2023

Serious adverse events – Idiopathic intracranial hypertension (IIH)

No algebraic manipulations or imputations were required. Incident rate ratio was calculated from reported IIH events and person-years. This study was not included in a meta-analysis.

Lahaije 2025

Menstruation

The proportion of youth menstruating at follow-up who were menstruating at baseline was calculated. The Freeman-Tukey double-arcsine variance stabilising transformation was applied prior to meta-analysis.

Lavender 2023

Gender dysphoria

The values reported as 'standard deviations' in the report, were actually standard errors. We therefore calculated the standard deviations from the confidence intervals and number of participants as follows:

$$SD = \sqrt{n} \times \left(\frac{CI_{upper} - CI_{lower}}{2 * invttail(n - 1{,}0.025)} \right)$$

The formulae in Section *Standardised mean difference* were applied to calculate the SMD and its SE for an uncontrolled before-after study.

Body image

The values reported as 'standard deviations' in the report, were actually standard errors. We therefore calculated the standard deviations from the confidence intervals and number of participants.

$$SD = \sqrt{n} \times \left(\frac{CI_{upper} - CI_{lower}}{2 * invttail(n - 1{,}0.025)} \right)$$

The formulae in Section *Standardised mean difference* were applied to calculate the SMD and its SE for an uncontrolled before-after study.

McGregor 2024

Social functioning

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

FOR PUBLIC CONSULTATION

Unadjusted and regression estimates were available. The estimate from the regression model was selected. The formulae in Section *Standardised mean difference* were then applied to calculate the partial SMD and its SE for a study with an independent control group.

Mental health – psychological functioning

Unadjusted and regression estimates were available. The estimate from the regression model was selected. The formulae in Section *Standardised mean difference* were then applied to calculate the partial SMD and its SE for a study with an independent control group.

Nunes-Moreno 2025

Mental health – suicidality diagnosis

No algebraic manipulations or imputations were required. This study reported a hazard ratio and confidence interval, which we presented in the review. This study was not included in a meta-analysis.

Schagen 2016

Menstruation

The proportion of youth menstruating at follow-up who were menstruating at baseline was calculated. The Freeman-Tukey double-arcsine variance stabilising transformation was applied prior to meta-analysis.

Turban 2020

Mental health – psychological functioning

Unadjusted and adjusted estimates were available from logistic regression models. We included the adjusted estimate. The adjusted odds ratio (and standard error) was converted to the SMD by dividing the log of the odds ratio by $\pi/\sqrt{3}$ (6).

van de Grift 2020

Pubertal development – Breast cup size

Analysed as a risk difference, separated by the Tanner stage 2/3 and 4/5. Events were defined as developing breast cup size >= B. The comparison group (no puberty suppressants) number of events and total number were halved as each intervention group (receipt of puberty suppressants at tanner stage 2/3, receipt of puberty suppressants at tanner stage 4/5) was compared against the same comparison group. In the case of odd numbers, a random number was used to choose which comparison group had the higher of the two numbers.

Pubertal development – Penile length

Analysed as a mean difference (cm), separated by the Tanner stage 2/3 and 4/5. The comparison group (no puberty suppressants) total number was halved as each intervention group (receipt of puberty suppressants at tanner stage 2/3, receipt of puberty suppressants at tanner stage 4/5) was compared against the same comparison group. In the case of odd numbers, a random number was used to choose which comparison group had the higher of the two numbers. The numbers for the groups were not explicitly reported. Therefore, missing values were calculated based on the proportions for similar outcomes (percent circumcised – 25% (24/96) Tanner group 2/3, 30% (29/96) Tanner group 4/5, 45% (43/96) Controls; and percent with descended testes – 24% (23/96) Tanner 2/3, 32% (31/96) Tanner 4/5, 44% (42/96) Control). The total sample size for the penile length groups was 103, for which, based on the similar proportions from percent circumcised and descended testes we used: 25% (25/103) in Tanner group 2/3, 32% (32/103) in Tanner group 4/5, 45% (46/103) in the comparison group.

Serious adverse events – Breast resection volume

FOR PUBLIC CONSULTATION

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

Analysed as a mean difference (cm), separated by the Tanner stage 2/3 and 4/5. The comparison group (no puberty suppressants) total number was halved as each intervention group (receipt of puberty suppressants at tanner stage 2/3, receipt of puberty suppressants at tanner stage 4/5) was compared against the same comparison group. In the case of odd numbers, a random number was used to choose which comparison group had the higher of the two numbers.

van der Miesen 2020

Functioning

No algebraic manipulations or imputations were required.

Mental health – psychological functioning

No algebraic manipulations or imputations were required.

Zaitoon 2025

Serious adverse effects – QtC interval

No algebraic manipulations or imputations were required.

Reported data was the mean change (calculated as baseline QTc interval minus post puberty suppressants QTc interval) and its confidence interval. The direction was reversed to represent mean change of post puberty suppressants minus baseline. This was achieved by multiplying the reported mean change and its confidence interval limits by -1.

FOR PUBLIC CONSULTATION

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

Supplement 1.D. Differences between the protocol and the review

Changes from the protocol and methods not implemented are coded as follows.

1. Rows without shading indicate changes to the planned methods before commencing the screening (as documented in Appendix S1.E) or a method not used because it was not required.
2. Rows shaded in blue indicate a change made after finalising the protocol. These changes are not expected to make a material difference to the review but are noted for transparency.
3. Rows shaded green indicate a change made after finalising the protocol to implement additional safeguards to ensure accuracy and completeness.

	Section	Planned method	Change	Details (text, rationale or both)
1	3.1.2 Types of participants	Inclusion of studies with participants up to 26 years	None	Based on feedback from the Committee and the Reference panel on the draft protocol, we revised the population criteria to clarify that participants could be any age at follow-up provided that they were under 18 years (on average) before commencing puberty suppression. This change was made ahead of commencing screening, so is document here for completeness, rather than indicating a change from protocol (see Appendix S1.E).
2	3.1.3 Types of interventions	Inclusion of non-GnRHa puberty suppressants (e.g. for menstrual suppression)	None	Based on feedback from the Committee and the Reference panel on the draft protocol, we revised the intervention criteria to limit to GnRHa puberty suppressants. This change was made ahead of commencing screening, so is document here for completeness, rather than indicating a change from protocol (see Appendix S1.E).
3	3.1.3 Types of interventions – comparisons	N/A	None	We edited the text describing eligible comparators for clarity. No material changes were made to the eligibility criteria for comparisons (and comparators) of interest.
4	3.1.4 Types of outcomes	N/A	None	Based on feedback on the draft protocol from the Committee and the Reference panel, we edited description of the outcomes for clarity, elaborated on the domains of interest and to: Edited text describing eligible outcomes for clarity Elaborated on the outcome domains, eligible measures and the rationale for inclusions/exclusions in the outcomes table No material changes were made to the eligibility of outcomes, measures or length of follow-up
5	3.1.4 Types of outcomes – selection of results (where a study reports multiple eligible results for a synthesis)	We reported that we would establish a hierarchy for selecting results by seeking targeted advice from members of the Committee and/or Reference panel with relevant expertise.	Method not used	We did not need to use this process because, where multiple results were available, the most relevant of the measures available was clear.
6	3.2.2 Database searches	Search terms Supplement, Appendix S1.A	Addition of terms	The following terms were added to the database search strategies (after the protocol was submitted to NHMRC on 13 February 2026): Population: birth assigned, assigned at birth Intervention: Amenorrhea/, hormon* suppression therapy, (menstrual adj (suppression or management or manipulation))

PUBLIC CONSULTATION

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

	Section	Planned method	Change	Details (text, rationale or both)
7	3.3.1 Selection of studies	We planned to pilot guidance for title and abstract screening	Method not used	We did not implement this step, in part because the first tranche of screening was of the full text studies listed by other SRs (Supplement 2). We took additional steps to ensure no studies were missed. • Two senior authors (SB and SMc) screened all studies. • We discussed screened studies at regular intervals to ensure a a consistent approach • We screened a high proportion of full text papers, retrieving papers that appeared to be ineligible, but which confirmation from full text was desired. • 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
8	3.3.2 Data extraction and management – study characteristics	We planned for a second author to extract study characteristics from a 10% sample of studies	Change to method	All study characteristics were extracted by a single author (SB), and uncertainties resolved with a second author (SM) or the senior biostatistician. All included studies were read, by multiple team members (SB, MP, DC, ST) as part of the evidence evaluation process. • MP assessed risk of bias of all studies and extracted relevant design features as part of those assessments. The assessments were discussed with JM and verified by SB • ST and DC extracted quantitative data from all studies. The data was checked by ST, discussed and verified by JM, and checked by SB when presented in forest plots This safeguarded against missing design features, misinterpretation of comparisons or outcomes etc. The quantitative data and risk of bias assessments were piloted by multiple authors, including the senior biostatistician (JM). There were extensive discussions of design features and results data among the team.
9	3.3.2 Data extraction and management – study characteristics	We reported that we would consider using extracted characteristics from other SRs	Method not used	We reviewed this data carefully and compared across SRs. We decided that the data did not cover the characteristics we planned to extract and were concerned about the categorisation of study designs (without explicit consideration of design features) and comparisons (i.e. unclear what co-interventions were received and by whom).
10	3.3.2 Data extraction and management – results data	We reported that 3 authors would pre-test the data extraction form	3 statisticians pre-tested	We decided that the complexity of the data and study designs were such, that three authors trained in biostatistics (JM, ST, DC) should independently test the data extract and coding of results data. This step included confirming that the comparison and outcomes had been appropriately coded by the lead author (SB) for the pilot studies.
11	3.3.2 Data extraction and management – results data	We reported that one author would extract data	Two authors extracted all results data	We decided that the complexity of the data and study designs were such, that two authors (ST and DC, both statisticians) should independently extract the results data. This step included confirming that the comparison and outcomes had been appropriately coded by the lead author (SB)
12	3.3.4 Measures of treatment effect	Our protocol addressed anticipated measures of	Added effect measures that arose from	We found studies that reported eligible results as: • count data, • time to event data, and

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

	Section	Planned method	Change	Details (text, rationale or both)
		treatment effect for continuous and dichotomous outcomes	data reported in included studies	• data that had an intuitive interpretation on the original scale (i.e. where reporting a mean difference was clearly preferable to an standardised mean difference) Effect measures for these data are now in the methods.
13	3.3.8 Assessment of biases due to missing results	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	Method not used	We were unable to apply this framework. • As is typical of non-randomised studies, only one of included studies had a registry entry or a protocol specific to the reported analyses • 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
14	3.3.8 Assessment of biases due to missing results - integrity	N/A	No change	Based on feedback from the Committee and the Reference panel on the draft protocol, we report our rationale for not performing integrity checks on the included studies. The INSPECT-SR tool is yet to be adapted for NRSIs. In the absence of a well-validate tool to assess integrity, and the low likelihood of finding information required for assessment, we did not feel that we could attest to the validity of checks using existing tools or that the checks would be informative.
15	3.3.9 Data synthesis – summary and synthesis when MA is not possible	We prespecified methods for reporting results and for considering alternative synthesis methods if meta-analysis could not be undertaken	Methods not used	For all outcomes, we were able to undertake meta-analysis or report an effect estimate from an individual study (if this was the only study available for the outcome).
16	3.3.9 Data synthesis – subgroup analysis	We prespecified subgroup analyses to examine whether age and sex assigned at birth explained difference in effects	Method not used	Findings were consistent across studies in the meta-analyses, so these analyses were not required. Some studies reported results (effect estimates) for subgroups, which we retained in our forest plots / reporting.
17	3.3.9 Data synthesis – sensitivity analysis	We planned 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	Method not used	There was little or no heterogeneity in any of the meta-analyses, so the analyses were not required.
18	3.3.10 Summary of findings tables and assessment of certainty of the evidence	We had reported that upgrading would be considered in the risk of bias assessment (as per an option with ROBINS-I)	Elaborated on method	We clarified the planned approach to upgrading.
19	3.3.10 Summary of findings tables and assessment of certainty of the evidence	We planned to “rate down for serious imprecision when the confidence interval crosses one threshold”	Elaborated on method	We clarified our approach “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.

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

	Section	Planned method	Change	Details (text, rationale or both)
20	3.3.11 Interpretation of findings	We did not address how we would determine 'large' effects	Elaborated on method	When we had evidence suggesting a large effect (sufficient to rate up certainty), we confirmed this interpretation with the Committee.

FOR PUBLIC CONSULTATION

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

Supplement 1.E. Response to the Methodological review and other feedback

Response to feedback on the protocol

In the table below we outline the main feedback received on the protocol and how it was addressed. This includes feedback from the independent methodological review contracted by the NHMRC (MR), the Gender Guideline Development Committee and the Gender Guideline Reference Panel.

The 'Change to methods' column identifies whether the comment led to a change in methods.

1. Blue rows indicate where a comment prompted a material change to methods, such as a change to the eligibility criteria, search strategy or synthesis method. These changes were made ahead of screening studies.
2. Other rows (marked as 'no change') indicate comments that prompted a change in reporting the protocol or the review itself. These changes include clarification of methods (e.g. definitions, rationale), revision of background, and adding supporting references. Some raised points we have addressed in the discussion.

Comments that required action are listed. Comments endorsing the prioritised outcomes or other aspects of the protocol are not listed). All comments have been deidentified as per requirements for NHMRC approved guidelines.

Section		Comment	Response / action	Change to methods
Methodological review				
MR1	METHODS - TYPES OF STUDIES	Query re. rationale for excluding conference abstracts	Added a rationale for not extracting data from conference abstracts which provide incomplete reporting of information, with an associated risk of bias	No change to methods
MR2	METHODS - SEARCH	Suggestion to have a PRESS appraisal of the search done	A PRESS appraisal was provided by an information specialist, experienced in developing and appraising searches for systematic reviews and with topic expertise. In response to the appraisal, the MEDline search strategy was revised. No additional included studies were retrieved. No changes were made to the other database strategies.	Revision to search prior to screening
MR3	METHODS - DATA COLLECTION	Query re. how unpublished data from trial registries would be handled if found	Edited text to state that this data would not be used (noting that no such data were found)	No change to methods
MR4	METHODS – POPULATION	Query re handling of population age, especially in studies with a mix of people above and below 18 years	See below re. GGDC and the GGRP	Change made ahead of screening
MR5	METHODS – OUTCOMES	Clarification required on length of follow-up	Length of follow-up has been clarified, also considering suggestions received from the GGDC and the GGRP	Change made ahead of screening
MR6	METHODS - OTHER SYNTHESIS	Query about the threshold to be used for vote counting (direction of effect) - null or a minimally important difference	Where an effect size is available, it is generally possible to calculate an effect estimate in the absence of a measure of variance. For this reason, we assumed the null as the basis for vote counting. In practice this method was not used.	No change to methods
MR7	METHODS – RISK OF BIAS / GRADE	Clarification on upgrading of evidence	Clarified the approach in the risk of bias and GRADE sections	Change made ahead of screening
Comments from the GGDC and GGRP				

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

	Section	Comment	Response / action	Change to methods
1	OVERARCHING	Comments on involvement of people with lived experience	We have elaborated on how those with lived experience and other interest holders contributed to the systematic review (but not the guideline development more generally)	No change to methods
2	BACKGROUND	Comments on background that fall outside the evidence synthesis	There is consensus guidance on what to report in a background to a systematic review[1], the purpose of which is to provide context for the synthesis of evidence about effects of a treatment. Where suggestions pertained to evidence related to GRADE factors not addressed in the systematic review (e.g. costs, acceptability, feasibility), these are not addressed in the background.	No change to protocol or review report
3	BACKGROUND	Comments on appropriateness of addressing health equity	Ensuring that evidence synthesis directly addresses health equity is a foundational principle of both the Cochrane collaboration and the GRADE working group. Both organisations provide guidance on methods for achieving this in evidence synthesis[2]. We strive to adhere to these principles and methods, and aim to address this in all aspects of the SR.	No change to protocol or review report
4	OBJECTIVES	Comments on addressing PS and GAH separately	We acknowledge the close interlinking of evidence on PS and GAH, and the conceptual and practical complexity of reviewing these therapies separately. Separate SRs were required to inform interim guidance on PS.	No change to methods
5	BACKGROUND	Comments on supporting evidence for the background	The background has been fully referenced to ensure that the underpinning evidence is provided.	No change to methods
6	BACKGROUND	Various comments on GnRHa therapies used in Australia, legal requirements etc	Changes made to background.	No change to methods
7	BACKGROUND/ METHODS	Suggestion to explain why another SR is needed, given that multiple exist	Paragraph added to background to explain limitations of using existing SRs, in the context of SRs on this topic	No changes to methods
8	POPULATION	The age cutoff was queried, with concerns raised that inclusion of people 26 years or under excludes the possibility of long-term follow up	The eligibility criteria were revised to • include studies in which puberty suppression was commenced before 18 years • remove any restrictions on age at follow-up • add explanation of how studies where some may have been older than 18 when commencing PS	Change to eligibility criteria made ahead of screening
9	INTERVENTION	Queries about inclusion of non-GnRHa medications, such as progestins and antiandrogens	The eligibility criteria were revised to exclude these medications	Change to eligibility criteria made ahead of screening

[1] PRISMA 2020 consensus statement on reporting systematic reviews doi: https://doi.org/10.1136/bmj.n71 ; Cochrane handbook Chapter III. Reporting the review https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-iii

[2] Cochrane handbook of interventions, Chapter 16: Equity and specific populations https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-16 .GRADE guidance on health equity (1st of 4 papers): https://pubmed.ncbi.nlm.nih.gov/28412464/

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

	Section	Comment	Response / action	Change to methods
10	OUTCOMES / OUTCOMES TABLE	Clarification required on length of follow-up	Suggestions considered and addressed in outcomes table	Change made ahead of screening
11	OUTCOMES / OUTCOMES TABLE	Suggestion to consider bone mineral density (BMD) / osteoporosis in studies that compare PS + GAH v GAH Other comments about bone health outcomes, clarifying fractures etc	This comparison warrants consideration and will be discussed with the GGDC for inclusion as part of the review of GAH. The close interlinking of evidence on PS and GAH adds considerable complexity to addressing this question – and others related to long-term effects - within the PS SR, especially for outcomes likely to be confounded by exposure to other interventions or factors over time. We have done the following. • Outcomes table. Clarified the measures eligible for 'fracture and functional outcomes, including criteria for paediatric osteoporosis for which fractures are the primary diagnostic criteria (not BMD). • Outcomes table. Added description of the limitations of BMD in adolescence as a surrogate for bone health. • Supplement 2. excluded studies. Included a filter to identify all studies that measured BMD and/or reported on fractures irrespective of whether they met the SR study design criteria. (to view studies, filter by 'Measures BMD, bone age, bone markers' = 1) • Discussion. Addressed bone health	No changes to methods
12	OUTCOMES / OUTCOMES TABLE	Multiple comments relating to - serious adverse events and how defined - consideration of harms, especially in relation to other prioritised outcomes (e.g. whether potential adverse effects on mental health would be considered)	Addressed as follows in the outcomes table • Provided a definition of 'serious adverse event' (SAE), referencing NHMRC and other major sources • Clarified the scope of outcomes eligible for consideration under the SAE domain, and the approach to identifying these events • Clarified how the SAE domain relates to other specified outcomes, including AEs of special interest (bone health, fertility) and outcomes identified in the logic model as primary and secondary goals of treatment	No change to methods
13	OUTCOMES / OUTCOMES TABLE	Comments on 'satisfaction with care' and suggestion to add "decision-making certainty/confidence in embodiment goals"	We considered any outcome related to this domain for inclusion. Outcomes framed as suggested would have been eligible since they aligns with the conceptualisation of satisfaction with the care process identified by Cochrane[3]. However, we did not find any eligible studies, so these outcomes are addressed in the discussion.	No change to methods

[3] Cochrane consumers and communication review group. Outcome taxonomy https://www.latrobe.edu.au/__data/assets/pdf_file/0012/234210/taxonomy-of-relevant-outcomes.pdf

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

FOR PUBLIC CONSULTATION

Response to feedback on the systematic review report

In the table below we outline the main feedback received from the independent methodological review contracted by the NHMRC (MR) on the systematic review report and how it was addressed.

Methodological review

Section		Comment	Response	Action
Methodological review				
MR1	METHODS – OUTCOME SELECTION	Results were selected 'based on relevance and availability of data', whereas the protocol specified a hierarchy developed with expert advice. Appendix S1.D does transparently note this as a deviation, and states that the hierarchy was not needed. However, **the choices as to what result was selected are not immediately evident.** For example, the total score of the Body Image Scale was selected for Carmichael 2021 and the secondary sex characteristics subscale for de Vries 2011 and Lavender 2023. It may be helpful to readers if the reviewers include a table or additional appendix that lists the eligible results for each study and outcome domain and the reason each result was selected.	In the CoIS table (supplement 2), we • list the outcomes for which results are available from each study by coded by outcome domain, • report which result was selected for each domain, and • indicate if data were incompletely report or some other factor precluded selection of an eligible result. It was not in our protocol to list all results (i.e. the data) and, in our view, to do so would risk of misinterpretation because results would be presented out of context (i.e. without calculation of effect estimates or accompanying appraisals and consideration of factors that influence certainty).	No change
MR2	METHODS – SEARCHING OTHER SOURCES	The NHMRC and the Committee were asked about studies not retrieved by the search. However, the report does not state whether any studies were identified this way; we suggest that this detail be added to the report.	Noted	We have added a sentence in Section 4.1 "No additional studies were identified by the Committee."
MR3	METHODS – ASSESSMENT OF RISK OF BIAS	The seven domains listed in the main text (page 18) follow ROBINS-I, whereas Supplement 3 uses the six domains (plus an overall judgement) of ROBINS-I V2. Both ROBINS-I and ROBINS-I V2 are cited, so it is unclear which version was used. It appears to be ROBINS-I V2, but this is not stated explicitly.	We wrote the protocol before receiving the updated version of ROBINS-I for some non-randomised study designs, so the protocol was based on version 1.	Text updated to version 2 in methods section 3.3.3

FOR PUBLIC CONSULTATION

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

	Section	Comment	Response	Action
MR4	METHODS – DATA SYNTHESIS	The report does not discuss the possible implications of skewed data. This may be relevant for the data from the Carmichael 2021 study. Appendix S1.C details that means and standard deviations for this study were estimated from medians and quartiles. With a sample size of 44, this study may or may not be considered small, which may be worth consideration by the NHMRC.	Carmichael is a before-after study, in which (for the numerator) we calculated the difference in means before and after puberty suppressants, which is equivalent to the mean of the individual differences. The distribution of the differences is likely to be less skewed than the distributions of the original observations (Vickers 2005 BMC Medical Research Methodology 2005, 5:35 doi:10.1186/1471-2288-5-35). This comment applies to all before-after studies. A further point specifically in relation to Carmichael was that we estimated the means and standard deviations from the reported 25th, 50th, and 75th percentiles. To do this, we used the formula proposed by Wan et al (BMC Medical Research Methodology 2014, 14:135) for samples larger than 30. These formulae work well for estimating the mean from the percentiles but can introduce a slight bias into the calculation of the standard deviation.	No change
MR5	METHODS - CERTAINTY	Section 3.3.10 notes that certainty was not rated down for imprecision where a confidence interval crossed a threshold only modestly (page 23). This is reasonable, but it is not described in the protocol or recorded in Appendix S1.D, and its application could be clarified. For example, the intervals for functioning (SMD: −0.30, 95%CI: −0.45 to −0.15) and psychological functioning (SMD −0.39, 95% CI −0.68 to −0.11) cross the threshold of −0.2 and were not rated down (footnotes r and t). The interval for gender dysphoria crosses the same threshold by a similar margin (lower limit −0.25) and was rated down (footnote g). We suggest the reviewers record the change in Appendix S1.D and briefly explain how it was applied, so that readers can follow the ratings for these outcomes	In section 3.3.1, Imprecision, our methods description was: "Where the extent to which the confidence interval crossed the threshold was modest (e.g. 2nd decimal place for an SMD), we did not rate down to avoid different certainty judgements for results that are not materially different." We agree that this does not address all scenarios in the three examples (described below). For the specific examples, gender dysphoria differs from the other two examples. • **Functioning** (SMD: 0.30 lower, 95% CI 0.45 lower to 0.15 lower). Both the upper and lower bound of the CI favour PS. Although the upper bound (SMD=0.15) crosses the threshold by 0.05 SD units, we do consider this effect meaningfully different from the threshold of an SMD of 0.2. • **Psych functioning** (SMD: 0.39 lower, 95% CI 0.68 lower to 0.11 lower). Both the upper and lower bound of the CI favour PS. Although the upper bound (SMD=0.11) crosses the threshold by 0.09 SD units, we do consider this effect meaningfully different from the threshold of an SMD of 0.2. • **Gender dysphoria** (SMD: 0.06 lower, 95% CI 0.25 lower to 0.13 higher). We rated down for imprecision because the upper and lower bounds of the 95% confidence interval are compatible with effects in the opposite direction; the lower bound favours no PS (-0.25) whereas the upper bound favours PS (SMD=0.13). In practice however, the decision to rate this result down for imprecision has no impact on the certainty because the very low certainty is due to extremely serious risk of bias.	We have edited our methods description in the report and noted the difference from the protocol in S1.D

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

	Section	Comment	Response	Action
MR6	RESULTS – SEARCH RESULTS	The PRISMA flow diagram in Figure 2 has a few inconsistencies, and it does not appear to reconcile with the numbers reported in the accompanying text. The reviewers have stated that 145 reports were assessed for eligibility, 18 of which were included (15 studies) and 127 excluded, totalling 145. However, the studies awaiting assessment (n=2) and ongoing study (n=1) are unaccounted for in this total. Either the box labelled 'Reports assessed for eligibility' needs to read 148 (to account for the three remaining studies), or their numbers need to be carried forward. As it stands, these studies are not accounted for in the calculations of this flow diagram.	We confirmed that these studies were counted in the exclusions instead of being carried forward.	The PRISMA flow diagram in Figure 2 has been corrected
MR7 a	RESULTS – Summary of findings Outcome: Body image and satisfaction	On pages 36–38, the text reports 125 participants for body image and satisfaction, whereas Table 4 reports 118. 126.	Number in SoF table is correct, text had total in study.	Number of participants report in text has been corrected
MR7 b	RESULTS – Summary of findings Outcome: breast cup size	The text also reports 81 participants for breast cup size, whereas Table 4 reports 126	Number in SoF table is correct, text contained an error.	Number of participants report in text has been corrected
MR7 c	RESULTS – Summary of findings Outcome: Quality of life	Regarding the quality of life outcome, the text refers to 'extremely serious imprecision', whereas footnote p of Table 4 refers to 'serious imprecision'. This does not change the certainty rating, as the evidence was already rated down by three levels for extremely serious risk of bias.	The rating down in the footnote was initially based on crossing one threshold. While rating down for 'serious' or 'extremely serious' imprecision would have no impact on the certainty of this result (which is already very low), we proposed to the Committee that the result was too imprecise to interpret with which they agreed. That is, the Committee agreed that the upper bound of the confidence interval (SMD=1.5) was compatible with very large benefit and the lower bound (SMD=0.03) indicated little to no difference. The footnote has now been updated accordingly.	The footnote has been updated to 'extremely serious imprecision' to reflect the final determination with additional explanation

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

	Section	Comment	Response	Action
MR7 d	RESULTS – Summary of findings Outcome: idiopathic intracranial hypertension (IIH)	[1] Regarding idiopathic intracranial hypertension (IIH), footnote l of Table 4 (page 34) rates the evidence down for serious imprecision but also states that 'Given that concern arises only when the incidence of IIH increases above the threshold, the observed imprecision is unimportant.' We understand that the interval crosses the threshold in the direction that favours puberty suppressants, but rating down for imprecision while describing it as unimportant appears contradictory, and we suggest the reviewers revise this wording. [2] The threshold for this outcome (20 more people with IIH per 100,000 person-years; page 33) is also far more conservative than those for other outcomes, for example 1 fewer or more person with a diagnosis of suicidality per 100 people over 12 months (page 33). We recognise this reflects the serious consequences of IIH (footnote j, page 34), but it would help to explain how this threshold was set relative to the others. [3] Finally, the summary on page 39 refers to both low and very low certainty in the same sentence: '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 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, **very low** certainty evidence)'. We suggest the reviewers check this against the rating in Table 4.	We address the 3 points in turn [numbering is ours] [1] We judged that rating down for imprecision was appropriate because the confidence interval is wide and compatible with different interpretations (i.e. crosses a threshold for an important effect): • The lower bound of the confidence interval is compatible with 'important benefit' (i.e. a reduction in IIH with puberty suppressants) • The upper bound is compatible with 'trivial' or 'unimportant' harm (i.e. the effect is below the threshold but favours no puberty suppressants). We described this imprecision as 'unimportant' because IIH is an adverse event compared to no intervention or usual care. We tend to be concerned only when there is potential for an important increase in an adverse event with treatment, which there is no indication of from this data on IIH. We agree that the wording could have been clearer. [2] The reason why the threshold for this outcome is set so low is because it is an adverse event (so a high value is placed on avoiding this outcome), whereas a reduction in suicidality a secondary goal of treatment (as per the logic model, Figure 1 in the main report). [3] We have corrected this transcription error. The reported certainty was correct in all other instances (e.g. summary of findings table, text preceding brackets).	[1] Footnote edited for clarity. 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). [2] Threshold reported in footer edited to read "… this is a rare *adverse* event *(not an outcome that is a goal of treatment)…"* [3] Transcription error corrected '~~very~~ low certainty'

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

	Section	Comment	Response	Action
MR7e	RESULTS – Summary of findings Outcome: pubertal development – penile length	Regarding pubertal development (assigned male at birth): penile length, the certainty of evidence is rated as moderate (mean difference 3.26 cm shorter, 95% CI 3.44 to 3.09 cm shorter; Table 4, page 30), despite rating down for both risk of bias and indirectness (footnotes a and b, page 34). The moderate risk of bias judgement for van de Grift 2020 is justified in Supplement 3. The evidence was then rated up for a large effect, with the direction of effect reversed from usual pubertal development (footnote a, page 34; page 45). Although this does not necessarily require a change, a moderate rating would usually be expected from a single non-randomised study at moderate risk of bias with no other concerns.	We fully agree with the methodologists' point that it is usual to be concerned about bias in non-randomised studies and that these concerns often lead to a rating of low or very low certainty evidence. This is reflected in our rating of other outcomes. However, pubertal development is an exception, as we describe in the results text under 'Study setting and design' "…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 the outcome penile length, our concerns about risk of bias were offset first because the direction of effect was opposite to the expected trajectory and second because of the large effect size. Any bias would need to be very large to shift the estimate below the threshold for an important difference. We explain this in the results text for this outcome: "While the evidence was downgraded for risk of bias, 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." We have edited the footnote to make this explanation more prominent.	Methods updated to provide a fuller description of the approach to upgrading. Footnote edited as per italics: 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 upgraded* for a large effect (+1).
MR7f	RESULTS – Summary of findings Outcome: pubertal development – menstruating	Similarly, menstruating is rated as high certainty (none of the participants were menstruating at follow-up across four studies; Table 4, page 30). Here, rating down by two levels for risk of bias is offset by rating up by two levels for a very large effect (footnote d, page 34). Although not incorrect, we suggest the reviewers explain more fully why these ratings are higher than would usually be expected for this body of evidence. The explanation could focus on the large effects in the direction opposite to usual pubertal	The rationale for the certainty rating of the outcome 'menstruating' is similar to that for the outcome 'penile length', except that none of the 143 participants that were menstruating at baseline were menstruating at follow-up. In line with classic examples where non-randomised studies provide high certainty evidence (e.g. the before-and-after test in the 1920's of insulin for reducing mortality among people with diabetes), the reversal of the usual direction of effects in all participants gives high confidence in the effects. We explain this in the results text for this outcome: "While all evidence from uncontrolled before-after 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,	Methods updated to provide a fuller description of the approach to updating. Footnote edited as per italics: 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

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

Section		Comment	Response	Action
		development, alongside a reflection on the GRADE considerations for rating up	including study limitations. For this reason, the evidence was upgraded to high certainty." We have edited the footnote to make this explanation more prominent.	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)

FOR PUBLIC CONSULTA

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

Supplement 2. Characteristics of studies (separate Excel file)

S2A. Characteristics of included studies [excel worksheet 1]

S2B. Studies assessed (including exclusions with reason) [excel worksheet 2]

Supplement 2.B reports decisions made when screening the full text of all studies from other major systematic reviews and from the updated search. The table reports the source of the study, the rationale for each decision and maps the decisions to those in other major systematic reviews.

- included studies
- excluded studies
- studies awaiting assessment
- ongoing studies

FOR PUBLIC CONSULTATION

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

Supplement 3. Risk of bias assessments

Becker-Hebly 2021

Setting: Germany (Hamberg Identity Service)

Design: Retrospective cohort study (RCS)

Confounders measured
Age at OR pubertal stage*: No Sex assigned at birth: No Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: YSR / ASR Total problem score, Table 3; Kid-screen-27/SF-8 – Mental dimension, Table 4

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Moderate**	(i) Confounding possible, although probably not substantial. Age: not included in regression analysis, but similar between groups. Sex assigned at birth: not included in regression analysis, but similar between groups. Social factors: No data provided on these factors. Baseline measure of the outcome: not included in regression analysis, but similar between groups. (ii) No important concerns about serious residual confounding. No concerns about the validity and reliability of measurement of confounding domains, no control for any post-intervention variables that could have been affected by the intervention, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls).
2. Bias in classification of interventions: **Low**	(i) No important concerns about immortal time bias. Receipt of PS not distinguishable at the time when follow-up commenced, but all or nearly all outcome events probably occurred after the intervention and comparator strategies could be distinguished. (ii) No important concerns about misclassification (differential & non-differential). Receipt of PS ascertained from participants and verified by clinicians. Unlikely that PS prescription would not be recorded by either party.
3. Bias in selection of participants into the study (or into the analysis): **Low**	(i) No important concerns about immortal time bias. After the start of the intervention, there was no period in which outcomes were not ascertained or events were excluded from the analysis. (ii) No important concerns about selection bias. The cohort was not selected based on participant characteristics observed after the start of the intervention.
4. Bias due to missing data: **Critical**	Important concerns about missing data. Only 75/204 eligible youth provided follow-up data. Intervention status missing from those with no follow-up, since this was the point at which their intervention status was obtained. Missingness of data may be due to intervention and outcome.

FOR PUBLIC CONSULTATION

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

Domain	Judgement
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed between intervention groups. (ii) Outcome assessors were the youth and they were aware of the intervention received, but knowledge of intervention status is unlikely to have influenced outcome assessment.
6. Bias in selection of the reported result: **Moderate**	Selective reporting possible. There was no protocol (or statistical analysis plan) identified from which to determine if measures or analyses reported were selected on the basis of results.
Overall risk of bias: **Critical**	Critical risk of bias due to missing data.

Risk-of-bias judgements: Outcome / result: CGAS Global functioning

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Moderate**	As above
2. Bias in classification of interventions: **Low**	As above
3. Bias in selection of participants into the study (or into the analysis): **Low**	As above
4. Bias due to missing data: **Critical**	As above
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed between intervention groups. (ii) Outcome assessors were aware of the intervention received by study participants, but knowledge of intervention status is unlikely to have influenced outcome assessment. Outcome collected during clinical assessments. Because assessment of the outcome took place in a visit unrelated to whether the youth had received PS, having received PS was unlikely to affect how they would respond.
6. Bias in selection of the reported result: **Moderate**	Selective reporting possible. There was no protocol (or statistical analysis plan) identified from which to determine if analyses reported were selected on the basis of results. Selective reporting on the basis of results from multiple outcome measures unlikely.
Overall risk of bias: **Critical**	Critical risk of bias due to missing data.

FOR PUBLIC CONSULTATION

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

Carmichael 2021

Setting: United Kingdom (Gender Identity Development Service)

Design: Uncontrolled before-after studies (including interrupted time series (ITS))

Confounders measured
Age at OR pubertal stage*: No Sex assigned at birth: No / not required for menstruation Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: Utrecht Gender dysphoria score, 12 mnths, Table 5; Body image scale - overall score, 12mths, Table 5; Side effects, text p12-13; Menstruation, 12mnths, Table 1*

Decide whether to proceed with a risk-of-bias assessment (screening): No, because the study compared a single pre-intervention measurement with a single post-intervention measurement such that it is impossible to distinguish the intervention effect from a time effect

Domain	**Judgement**
1. Bias due to confounding: **Not assessed**	Not assessed.
2. Bias in selection of participants into the study (or into the analysis): **Not assessed**	Not assessed.
3. Bias in selection of pre- versus post-intervention time points: **Not assessed**	Not assessed.
4. Bias due to missing data: **Not assessed**	Not assessed.
5. Bias in measurement of the outcome: **Not assessed**	Not assessed.
6. Bias in selection of the reported result: **Not assessed**	Not assessed.
Overall risk of bias: **Critical**	Critical risk of bias due to confounding.

FOR PUBLIC CONSULTATION

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

Costa 2015

Setting: United Kingdom (Gender Identity Development Service)

Design: Prospective cohort study (RCS)

Confounders measured
Age at OR pubertal stage*: No Sex assigned at birth: No Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: CGAS, Table 2, T0 and T3

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Moderate**	(i) Confounding possible, although probably not substantial. Age: not adjusted for, and not reported whether it differed between groups. Sex assigned at birth: not adjusted for, but ratio of assigned sex was similar between groups. Social factors: not adjusted for, and not reported whether any factor differed between groups. Baseline measure of the outcome: Not adjusted for, but baseline CGAS scores were similar between groups. (ii) No important concerns about serious residual confounding. No concerns about the validity and reliability of measurement of confounding domains, no control for any post-intervention variables that could have been affected by the intervention, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls).
2. Bias in classification of interventions: **Low**	(i) No important concerns about immortal time bias. Receipt of PS distinguishable at the time when follow-up commenced. (ii) No important concerns about misclassification (differential & non-differential). Receipt of PS ascertained from patient records. Unlikely that PS prescription would not be recorded.
3. Bias in selection of participants into the study (or into the analysis): **Low**	(i) No important concerns about immortal time bias. After the start of the intervention, there was no period in which outcomes were not ascertained or events were excluded from the analysis. (ii) No important concerns about selection bias. The cohort was not selected based on participant characteristics observed after the start of the intervention.
4. Bias due to missing data: **Serious**	Important concerns about missing data. Only 71/201 (35%) eligible youth provided follow-up data at T3 (18 months from baseline). The number of participants providing data at this time point was similar, but no reasons for missing data were provided, so it is unclear if missing data (in intervention, confounders or the outcome) are likely to be related to the true value of the outcome.
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed between intervention groups.

FOR PUBLIC CONSULTATION

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

Domain	Judgement
	(ii) Whether outcome assessors were aware of the intervention received by participants is unclear; however, knowledge of the intervention received is unlikely to have influenced outcome assessment. Because assessment of the outcome took place in a visit unrelated to whether the youth had received PS, having received PS was unlikely to affect how clinicians would assess outcomes.
6. Bias in selection of the reported result: **Moderate**	Selective reporting possible. There was no protocol (or statistical analysis plan) identified from which to determine if analyses reported were selected on the basis of results.
Overall risk of bias: **Serious**	Serious risk of bias due to missing data.

FOR PUBLIC CONSULTATION

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

de Vries 2011

Setting: The Netherlands (Amsterdam gender identity clinic of the VU university medical center)

Design: Uncontrolled before-after studies (including interrupted time series (ITS))

Confounders measured
Age at OR pubertal stage*: No Sex assigned at birth: No Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: Body image (BIS) Secondary sex characteristics, Table 3 (all for INT1/COM1, natal female for INT2/COM2, natal male for INT3/COM3); Gender dysphoria scale (UGDS), Table 5, (all for INT1/COM1, natal female for INT2/COM2, natal male for INT3/COM3

Decide whether to proceed with a risk-of-bias assessment (screening): No, because the study compared a single pre-intervention measurement with a single post-intervention measurement such that it is impossible to distinguish the intervention effect from a time effect

Domain	**Judgement**
1. Bias due to confounding: **Not assessed**	Not assessed.
2. Bias in selection of participants into the study (or into the analysis): **Not assessed**	Not assessed.
3. Bias in selection of pre- versus post-intervention time points: **Not assessed**	Not assessed.
4. Bias due to missing data: **Not assessed**	Not assessed.
5. Bias in measurement of the outcome: **Not assessed**	Not assessed.
6. Bias in selection of the reported result: **Not assessed**	Not assessed.
Overall risk of bias: **Critical**	Critical risk of bias due to confounding.

FOR PUBLIC CONSULTATION

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

Fisher 2024

Setting: Italy (Florence Gender Clinic)

Design: Uncontrolled before-after studies (including interrupted time series (ITS))

Confounders measured
Age at OR pubertal stage*: No
Sex assigned at birth: No / not required for menstruation
Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No
Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: Menstruation text, page 5, col2; Tanner stages, Figure 2 (results for Tanner stage could not be used for meta-analysis or to calculate effect estimates).

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Low (except for concerns about uncontrolled confounding)**	(i) Confounding not expected. Confounding factors are unlikely to explain the halt in puberty, so adjustment would not be needed. (ii) No important concerns about serious residual confounding. No concerns about the validity and reliability of measurement of confounding domains, no control for any post-intervention variables that could have been affected by the intervention, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls).
2. Bias in selection of participants into the study (or into the analysis): **Low**	The study prospectively follows a specific group of units right through from start (pre-intervention) to finish (post-intervention).
3. Bias in selection of pre- versus post-intervention time points: **Low**	(i) Pre-intervention measurements were unlikely to be affected by any preparatory phases of the intervention. (ii) Specification of the distinction between pre-intervention time points and post-intervention time points is unlikely to have been influenced by the outcome data.
4. Bias due to missing data: **Moderate**	Possible concerns about missing data. Only 36/54 (67%) eligible youth provided follow-up data (18 dropped out prior to starting GnHRa). Missingness of data may be related to the true value of the outcome, but not strongly, given that baseline mental health and psychosocial functioning were similar between those who completed follow-up and those who dropped out.
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed before and after the intervention. (ii) Outcome assessors were aware of whether the intervention had happened, but knowledge of whether the intervention had happened is unlikely to have influenced outcome assessment. Participants attended an endocrinological assessment every three months, during which a physical examination (including Tanner stage assessment) was performed. Because assessment of the outcome took place in a visit

FOR PUBLIC CONSULTATION

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

Domain	Judgement
	unrelated to whether the youth had received PS, having received PS was unlikely to affect how clinicians would assess Tanner stage.
6. Bias in selection of the reported result: **Moderate**	Selective reporting possible. There was no protocol (or statistical analysis plan) identified from which to determine if analyses reported were selected on the basis of results.
Overall risk of bias: **Moderate**	Moderate risk of bias due to missing data and selection of the reported result.

FOR PUBLIC CONSULTATION

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

Karamanis 2023

Setting: Sweden (various national registers)

Design: Retrospective cohort study (RCS)

Confounders measured
Age at OR pubertal stage*: No Sex assigned at birth: No Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: IIH incidence rate per 10000 person years, %

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Moderate**	(i) Confounding possible, although probably not substantial. Age: not adjusted for, and age at entry date differed between groups (17.8 years versus 23.1 years). Sex assigned at birth: not adjusted for, but did not differ substantially between groups. Social factors: no mention of distribution of social factors, but these are unlikely to have influenced the outcome. Baseline measure of the outcome: Not applicable as the outcome is incidence of a serious adverse event (which cannot have occurred at baseline). (ii) No important concerns about serious residual confounding. No concerns about the validity and reliability of measurement of confounding domains, no control for any post-intervention variables that could have been affected by the intervention, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls).
2. Bias in classification of interventions: **Low**	(i) No important concerns about immortal time bias. Receipt of PS distinguishable at the time when follow-up commenced. (ii) No important concerns about misclassification (differential & non-differential). Receipt of PS ascertained from National Prescribed Drug Register. Unlikely that PS prescription would not be recorded.
3. Bias in selection of participants into the study (or into the analysis): **Low**	(i) No important concerns about immortal time bias. After the start of the intervention, there was no period in which outcomes were not ascertained or events were excluded from the analysis. (ii) Selection bias is unlikely. Follow-up started at first GD diagnosis or GnRHa initiation, hence was based only on characteristics observed before the start of the intervention.
4. Bias due to missing data: **Low**	No important concerns about missing data. Diagnosis of idiopathic intracranial hypertension ascertained from the National Patient Register. Receipt of PS ascertained from the National Prescribed Drug Register. Unlikely that diagnosis or PS prescription would not be recorded in each register. No missing data for confounding variables.
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed between intervention groups.

FOR PUBLIC CONSULTATION

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

Domain	Judgement
	(ii) Whether outcome assessors were aware of the intervention received by participants is unclear; however, knowledge of the intervention received is unlikely to have influenced outcome assessment. Diagnosis of idiopathic intracranial hypertension was ascertained from the National Patient Register. Because assessment of the outcome took place in a visit unrelated to whether the youth had received PS, having received PS was unlikely to affect how clinicians would diagnose this condition.
6. Bias in selection of the reported result: **Moderate**	Selective reporting possible. There was no protocol (or statistical analysis plan) identified from which to determine if analyses reported were selected on the basis of results.
Overall risk of bias: **Moderate**	Moderate risk of bias due to confounding and selection of the reported result.

FOR PUBLIC CONSULTATION

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

Lahaije 2025

Setting: The Netherlands (Radboud UMC Expert Center for Sex & Gender)

Design: Uncontrolled before-after studies (including interrupted time series (ITS))

Confounders measured
Age at OR pubertal stage*: Yes (results reported by Tanner stage; only data for Tanner 4/5 eligible because none of those at Tanner 2/3 had reached menarche before commencing PS)
Sex assigned at birth: No / not required for menstruation
Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No
Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: menstruation [for those at menarche at start of PS; 6 months from commencement of PS, Text p4, column 2

Decide whether to proceed with a risk-of-bias assessment (screening): No, because the study compared a single pre-intervention measurement with a single post-intervention measurement such that it is impossible to distinguish the intervention effect from a time effect

Domain	**Judgement**
1. Bias due to confounding: **Not assessed**	Not assessed.
2. Bias in selection of participants into the study (or into the analysis): **Not assessed**	Not assessed.
3. Bias in selection of pre- versus post-intervention time points: **Not assessed**	Not assessed.
4. Bias due to missing data: **Not assessed**	Not assessed.
5. Bias in measurement of the outcome: **Not assessed**	Not assessed.
6. Bias in selection of the reported result: **Not assessed**	Not assessed.
Overall risk of bias: **Critical**	Critical risk of bias due to confounding.

FOR PUBLIC CONSULTATION

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

Lavender 2023

Setting: United Kingdom (Gender Identity Development Service)

Design: Uncontrolled before-after studies (including interrupted time series (ITS))

Confounders measured
Age at OR pubertal stage*: No
Sex assigned at birth: No
Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No
Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: Utrecht Gender dysphoria overall score, Table 3 (COM. baseline; INT: 1 year after GnRHA); Body image scale - secondary sex characteristics, Table 3, COM baseline, INT 1 year after GnRHA

Decide whether to proceed with a risk-of-bias assessment (screening): No, because the study compared a single pre-intervention measurement with a single post-intervention measurement such that it is impossible to distinguish the intervention effect from a time effect

Domain	**Judgement**
1. Bias due to confounding: **Not assessed**	Not assessed.
2. Bias in selection of participants into the study (or into the analysis): **Not assessed**	Not assessed.
3. Bias in selection of pre- versus post-intervention time points: **Not assessed**	Not assessed.
4. Bias due to missing data: **Not assessed**	Not assessed.
5. Bias in measurement of the outcome: **Not assessed**	Not assessed.
6. Bias in selection of the reported result: **Not assessed**	Not assessed.
Overall risk of bias: **Critical risk of bias**	Critical risk of bias due to confounding.

FOR PUBLIC CONSULTATION

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

McGregor 2024

Setting: USA

Design: Retrospective cohort study (RCS)

Confounders measured
Age at OR pubertal stage*: No Sex assigned at birth: Affirmed gender (male, female, non-binary) Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: YSR Total problem score / -4.5 Table 2

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Moderate**	(i) Confounding possible, although probably not substantial. Age: not included in regression analysis, but when the analysis was restricted by age, the results were not importantly affected (as stated by authors, results not provided). Sex assigned at birth: included correlated variable (affirmed gender). Social factors: distribution of ethnicity not importantly different across groups (most participants white); in addition, participants noted to be predominantly from the New England area, which is often considered a liberal area of the USA. (ii) No important concerns about serious residual confounding. No concerns about the validity and reliability of measurement of confounding domains, no control for any post-intervention variables that could have been affected by the intervention, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls).
2. Bias in classification of interventions: **Low**	(i) No important concerns about immortal time bias. Receipt of PS distinguishable at the time when follow-up commenced. (ii) No important concerns about misclassification (differential & non-differential). Receipt of PS ascertained from chart review. Unlikely that PS prescription would not be recorded.
3. Bias in selection of participants into the study (or into the analysis): **Moderate**	(i) No important concerns about immortal time bias. After the start of the intervention, there was no period in which outcomes were not ascertained or events were excluded from the analysis. (ii) Selection bias possible. Youth with psychosocial problems interfering with assessment or treatment may not be prescribed PS and, subsequently, may not have attended for an assessment of GAH because, for example, their mental health was poorer or they sought care from another clinic.
4. Bias due to missing data: **Low**	No important concerns about missing data. Youth required to complete measurement instruments as part of assessment for GAH. Receipt of PS ascertained from chart review. Unlikely that PS prescription would not be recorded. No missing data for confounding variable (affirmed gender).

FOR PUBLIC CONSULTATION

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

Domain	Judgement
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed between intervention groups. (ii) Outcome assessors were aware of the intervention received by study participants, but knowledge of intervention status is unlikely to have influenced outcome assessment. Outcome collected during routine clinical assessments for GAH. Because assessment of the outcome took place in a visit unrelated to whether the youth had received PS, having received PS was unlikely to affect how they would respond.
6. Bias in selection of the reported result: **Moderate**	Selective reporting possible. There was no protocol (or statistical analysis plan) identified from which to determine if analyses reported were selected on the basis of results. Selective reporting on the basis of results from multiple outcome measures unlikely due to the results of all subdomains of the YSR being presented.
Overall risk of bias: **Moderate**	Moderate risk of bias due to confounding, selection of participants into the study, and selection of the reported result.

Risk-of-bias judgements: Outcome / result: YSR scale Social problems score / -0.7 Table 2

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Moderate**	Confounding bias: Possible, and potentially substantial. No adjustment for confounding factors in the analysis. As per YSR Total problem score, age and social factors are accounted for; however, sex assigned at birth is not, and the distribution of gender differs across those receiving and not receiving PS. (ii) No important concerns about serious residual confounding. No concerns about the validity and reliability of measurement of confounding domains, no control for any post-intervention variables that could have been affected by the intervention, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls).
2. Bias in classification of interventions: **Low**	As above
3. Bias in selection of participants into the study (or into the analysis): **Moderate**	As above
4. Bias due to missing data: **Low**	As above
5. Bias in measurement of the outcome: **Low**	As above
6. Bias in selection of the reported result: **Moderate**	As above
Overall risk of bias: **Moderate**	Moderate risk of bias due to confounding, selection of participants into the study, and selection of the reported result.

FOR PUBLIC CONSULTATION

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

Nunes-Moreno 2025

Setting: USA (electronic health records in PEDSnet, a pediatric learning health system)

Design: Retrospective cohort study (RCS)

Confounders measured
Age at OR pubertal stage*: Age at first diagnosis of gender dysphoria Sex assigned at birth: Sex listed in the EHR, which most often reflected the sex assigned at birth Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): Race (White, Black, Asian, Other, Unknown), ethnicity (Hispanic, non-Hispanic, unknown), health insurance (public, private, other) Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: Diagnosis of suicidality in an ED or inpatient encounter, text p22, col2 (also Fig1, panel B)

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Serious**	(i) No important concerns about critical confounding factors. Age: included correlated variable (age at first diagnosis of gender dysphoria) in regression analysis. Sex assigned at birth: included in regression analysis. Social factors: included race, ethnicity, and health insurance in regression analysis. Baseline measure of the outcome: Not applicable as the outcome is diagnosis of suicidality (which cannot have occurred at baseline). (ii) No important concerns about the validity and reliability of measurement of confounding domains, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls). (iii) Concerns about controlling for a post-intervention variable that could have been affected by the intervention. Whether the patient was additionally prescribed a gender-affirming hormone therapy at least once during the study period was included in the regression analysis.
2. Bias in classification of interventions: **Low**	(i) No important concerns about immortal time bias. Receipt of PS distinguishable at the time when follow-up commenced. (ii) No important concerns about misclassification (differential & non-differential). Receipt of PS ascertained from EHR. Unlikely that PS prescription would not be recorded.
3. Bias in selection of participants into the study (or into the analysis): **Low**	(i) No important concerns about immortal time bias. After the start of the intervention, there was no period in which outcomes were not ascertained or events were excluded from the analysis. (ii) No important concerns about selection bias. The cohort was not selected based on participant characteristics after the start of the intervention.
4. Bias due to missing data: **Low**	No important concerns about missing data. It is possible that some data on PS prescription and suicidality are missing if youth attended sites not included in the study. No reason to suspect that missing data would be related to the outcome. There is

FOR PUBLIC CONSULTATION

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

Domain	Judgement
	some missing data for confounders (race, ethnicity), although youth with missing confounders are still included in the analysis.
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed between intervention groups. (ii) Whether outcome assessors were aware of the intervention received by participants is unclear; however, knowledge of the intervention received is unlikely to have influenced outcome assessment. Diagnosis of suicidality was ascertained from the EHR. Because assessment of the outcome took place in a visit unrelated to whether the youth had received PS, having received PS was unlikely to affect how clinicians would diagnose suicidality.
6. Bias in selection of the reported result: **Moderate**	Selective reporting possible. There was no protocol (or statistical analysis plan) identified from which to determine if analyses reported were selected on the basis of results.
Overall risk of bias: **Serious**	Serious risk of bias due to confounding.

FOR PUBLIC CONSULTATION

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

Schagen 2016

Setting: The Nethlerands (VU University Medical Center)

Design: Uncontrolled before-after studies (including interrupted time series (ITS))

Confounders measured
Age at OR pubertal stage*: No Sex assigned at birth: Not required (pubertal development results by sex assigned at birth) Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: Menstruation, 12 months, text p1128, col 1; Testicular volume, 12 months, text p1128, col 1 & Table 1.

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Low (except for concerns about uncontrolled confounding)**	(i) Confounding not expected. Confounding factors are unlikely to explain the halt in puberty, so adjustment would not be needed. (ii) No important concerns about serious residual confounding. No concerns about the validity and reliability of measurement of confounding domains, no control for any post-intervention variables that could have been affected by the intervention, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls).
2. Bias in selection of participants into the study (or into the analysis): **Low**	The study prospectively follows a specific group of units right through from start (pre-intervention) to finish (post-intervention).
3. Bias in selection of pre- versus post-intervention time points: **Low**	(i) Pre-intervention measurements were unlikely to be affected by any preparatory phases of the intervention. (ii) Specification of the distinction between pre-intervention time points and post-intervention time points is unlikely to have been influenced by the outcome data.
4. Bias due to missing data: **Low**	No important concerns about missing data. Forty-nine of 55 (89%) male-to-female participants provided baseline and follow-up data on testicular volume.
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed before and after the intervention. (ii) Outcome assessors were likely aware of whether the intervention had happened, but knowledge of whether the intervention had happened is unlikely to have influenced outcome assessment. Participants attended the clinic every three months, during which assessment of testicular volume (using a Prader orchidometer) was performed. Because assessment of the outcome took place in a visit unrelated to whether the youth had received PS, having received PS was unlikely to affect how clinicians would assess testicular volume.

FOR PUBLIC CONSULTATION

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

Domain	Judgement
6. Bias in selection of the reported result: **Low**	Selective reporting is unlikely. There was no protocol (or statistical analysis plan) identified, but no compelling reason why the numerical result being assessed is likely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. definitions), multiple measures of intervention effect, multiple methods for analysing the data, or multiple subgroups.
Overall risk of bias: **Low (except for concerns about uncontrolled confounding)**	Low risk of bias (except for concerns about uncontrolled confounding).

FOR PUBLIC CONSULTATION

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

van de Grift 2020

Setting: The Netherlands (single center using registry and patient record data)

Design: Retrospective cohort study (RCS)

Confounders measured
Age at OR pubertal stage*: No Sex assigned at birth: No / not required because results are reported and included in meta-analysis by gender (trans women, trans men) Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: Physical characteristics, transgender men, breasts cup, Table 3; Physical characteristics, transgender women, penile length, Table 3; Physical characteristics/surgery, transgender men, Mastectomy - total wt resected (g) [Table 3]/ total breast volume resected [Table 4]; Physical characteristics/surgery, transgender women, intestinal vaginoplasty [Tables 3 & 4]

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Moderate**	(i) Confounding possible, although probably not substantial. Age: not adjusted for, and age at follow-up and age at initial assessment differed between groups (for the total sample; data for age not reported separately by sex assigned at birth or gender). Sex assigned at birth: no adjustment necessary as data were analysed separately for trans men and trans women. Social factors: education differed between groups (for the total sample; data not reported separately by SAB), but this and other social factors are unlikely to have influenced the outcome. Baseline measure of the outcome: Not applicable as the outcome was not measured at baseline. (ii) No important concerns about serious residual confounding. No concerns about the validity and reliability of measurement of confounding domains, no control for any post-intervention variables that could have been affected by the intervention, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls).
2. Bias in classification of interventions: **Low**	(i) No important concerns about immortal time bias. Receipt of PS distinguishable at the time when follow-up commenced. (ii) No important concerns about misclassification (differential & non-differential). Receipt of PS ascertained from patient records. Unlikely that PS prescription would not be recorded.
3. Bias in selection of participants into the study (or into the analysis): **Moderate**	(i) No important concerns about immortal time bias. After the start of the intervention, there was no period in which outcomes were not ascertained or events were excluded from the analysis. (ii) Selection bias possible. Participants were included only if they were at least 18 at the point of data collection. Youth with psychosocial problems interfering with assessment or treatment may not be prescribed PS and, subsequently, may not have reattended the clinic because, for example, their mental health was poorer or they sought care from another clinic.

FOR PUBLIC CONSULTATION

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

Domain	Judgement
4. Bias due to missing data: **Low**	No important concerns about missing data. Only 10 of 210 (5%) eligible participants were excluded from the analysis.
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed between intervention groups. (ii) Whether outcome assessors were aware of the intervention received by participants is unclear; however, knowledge of the intervention received is unlikely to have influenced outcome assessment. Because assessment of the outcome took place in a visit unrelated to whether the youth had received PS, having received PS was unlikely to affect how clinicians would assess outcomes.
6. Bias in selection of the reported result: **Moderate**	Selective reporting possible. There was no protocol (or statistical analysis plan) identified from which to determine if analyses reported were selected on the basis of results.
Overall risk of bias: **Moderate risk of bias**	Moderate risk of bias due to confounding, selection of participants into the study, and selection of the reported result.

FOR PUBLIC CONSULTATION

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

van der Miesen 2020

Setting: The Nethlerands (VU University Medical Center)

Design: Retrospective cohort study (RCS)

Confounders measured
Age at OR pubertal stage*: Age when first attending the gender identity service or age when about to start gender-affirming hormones
Sex assigned at birth: Sex assigned at birth
Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): Ethnicity (Dutch, non-Dutch, unknown), level of education (prevocational education, preparatory scholarly education, higher general continued education), parent's marital status (both parents, other, unknown)
Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: Youth self report - internalizing, Table 2 (all); Youth self report - peer relations, Table 2 (all)

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Serious**	(i) Confounding possible, and potentially substantial. Age: not adjusted for, and the distribution of age differs across those receiving and not receiving PS. Sex assigned at birth: not adjusted for, but authors state that a multivariate analysis including assigned gender showed similar findings (results not provided). Social factors: no adjustment for ethnicity, level of education and parent's marital status, but these did not differ importantly between groups. Baseline measure of the outcome: Not applicable as the outcome was not measured at baseline. (ii) No important concerns about serious residual confounding. No concerns about the validity and reliability of measurement of confounding domains, no control for any post-intervention variables that could have been affected by the intervention, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls).
2. Bias in classification of interventions: **Low**	(i) No important concerns about immortal time bias. Receipt of PS distinguishable at the time when follow-up commenced. (ii) No important concerns about misclassification (differential & non-differential). Receipt of PS ascertained from patient records. Unlikely that PS prescription would not be recorded.
3. Bias in selection of participants into the study (or into the analysis): **Moderate**	(i) No important concerns about immortal time bias. After the start of the intervention, there was no period in which outcomes were not ascertained or events were excluded from the analysis. (ii) Selection bias possible. Participants in the PS group were included only if they were attending the clinic to start gender-affirming hormone treatment. Youth with psychosocial problems interfering with assessment or treatment may not be prescribed PS and, subsequently, may not have reattended the clinic because, for example, their mental health was poorer or they sought care from another clinic.

FOR PUBLIC CONSULTATION

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

Domain	Judgement
4. Bias due to missing data: **Low**	No important concerns about missing data. Only one of 451 (0.2%) eligible participants were excluded from the analysis for not having completed the questionnaire.
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed between intervention groups. (ii) Outcome assessors were the youth and they were aware of the intervention received, but knowledge of intervention status is unlikely to have influenced outcome assessment.
6. Bias in selection of the reported result: **Moderate**	Selective reporting possible. There was no protocol (or statistical analysis plan) identified from which to determine if analyses reported were selected on the basis of results.
Overall risk of bias: **Serious risk of bias**	Serious risk of bias due to confounding.

Risk-of-bias judgements: Outcome / result: Youth self report - internalizing, Table 3 (by sex assigned at birth); Youth self report - peer relations, Table 3 (by sex assigned at birth)

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Moderate**	(i) Confounding possible, although probably not substantial. Age: not adjusted for, but authors stated that a multivariate analysis including age, ethnicity, level of education and parent's marital status showed similar findings (results not provided). Sex assigned at birth: no adjustment necessary as data were analysed separately for men and women. Social factors: not adjusted for, but authors stated that a multivariate analysis including age, ethnicity, level of education and parent's marital status showed similar findings (results not provided). Baseline measure of the outcome: Not applicable as the outcome was not measured at baseline. (ii) No important concerns about serious residual confounding. No concerns about the validity and reliability of measurement of confounding domains, no control for any post-intervention variables that could have been affected by the intervention, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls).
2. Bias in classification of interventions: **Low**	As above.
3. Bias in selection of participants into the study (or into the analysis): **Moderate**	As above.
4. Bias due to missing data: **Low**	As above.
5. Bias in measurement of the outcome: **Low**	As above.

FOR PUBLIC CONSULTATION

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

Domain	Judgement
6. Bias in selection of the reported result: **Moderate**	As above.
Overall risk of bias: **Moderate**	Moderate risk of bias due to confounding, selection of participants into the study, and selection of the reported result.

FOR PUBLIC CONSULTATION

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

Turban 2020

Setting: United States (cross-sectional survey of a community-recruited sample of transgender adults)

Design: Retrospective cohort (Cross-sectional data collection; PS confirmed as received before outcome measured)

Confounders measured
Age at OR pubertal stage*: Current age Sex assigned at birth: Gender identity Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): Level of family support for gender identity, sexual orientation, education level, employment status, total household income, relationship status Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: Mental health - Past-month severe psychological distress, K6 ≥13, Tables 2 and 3

Decide whether to proceed with a risk-of-bias assessment (screening): Yes

Domain	Judgement
1. Bias due to confounding: **Serious**	(i) Confounding by critical factors possible, although probably not substantial. Age: included correlated variable (current age) in regression analysis. Sex assigned at birth: included correlated variable (gender identity) in regression analysis. Social factors: no social factors at baseline were included in the regression analysis. Baseline measure of the outcome: Not applicable as the outcome was not measured at baseline. (ii) No important concerns about the validity and reliability of measurement of confounding domains, and no evidence of serious uncontrolled confounding (e.g., from use of negative controls). (iii) Important concerns about controlling for several post-intervention variables that could have been affected by the intervention: current level of family support for gender identity, current sexual orientation, highest education level, current employment status, current total household income, current relationship status.
2. Bias in classification of interventions: **Low**	(i) No important concerns about immortal time bias. Receipt of PS distinguishable at the time when follow-up commenced. (ii) No important concerns about misclassification (differential & non-differential). Receipt of PS ascertained via participant self-report, as part of a survey addressing various aspects. Unlikely that participants would incorrectly report whether they had received PS or not.
3. Bias in selection of participants into the study (or into the analysis): **Low**	(i) No important concerns about immortal time bias. After the start of the intervention, there was no period in which outcomes were not ascertained or events were excluded from the analysis. (ii) No important concerns about selection bias. The cohort was not selected based on participant characteristics observed after the start of the intervention.
4. Bias due to missing data: **Moderate**	No important concerns about missing data on intervention status. Unclear if there was any missing data on the outcome or confounding variables (although it is unlikely that all participants completed every question in the survey).

FOR PUBLIC CONSULTATION

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

Domain	Judgement
5. Bias in measurement of the outcome: **Low**	No important concerns about measurement of the outcome because: (i) Measurement or ascertainment of the outcome is unlikely to have differed between intervention groups. (ii) Outcome assessors were the transgender adults and they were aware of the intervention received, but knowledge of intervention status is unlikely to have influenced outcome assessment.
6. Bias in selection of the reported result: **Moderate**	Selective reporting possible. There was no protocol (or statistical analysis plan) identified from which to determine if analyses reported were selected on the basis of results.
Overall risk of bias: **Serious**	Serious risk of bias due to confounding.

FOR PUBLIC CONSULTATION

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

Zaitoon 2025

Setting: Israel (Children and Adolescents Gender Clinic)

Design: Uncontrolled before-after studies (including interrupted time series (ITS))

Confounders measured
Age at OR pubertal stage*: No Sex assigned at birth: No Social factors (rurality; socio-economic status; supportive parent or caregiver or other; gender minority stress; ethnicity): No Baseline measure of the outcome: No

* at initiation of puberty suppression

Risk-of-bias judgements: Outcome / result: QTc interval, ≥ 2 months, 'On GnRH', Table 1 (trans females), Table 3 (transmales)

Decide whether to proceed with a risk-of-bias assessment (screening): No, because the study compared a single pre-intervention measurement with a single post-intervention measurement such that it is impossible to distinguish the intervention effect from a time effect

Domain	**Judgement**
1. Bias due to confounding: **Not assessed**	Not assessed.
2. Bias in selection of participants into the study (or into the analysis): **Not assessed**	Not assessed.
3. Bias in selection of pre- versus post-intervention time points: **Not assessed**	Not assessed.
4. Bias due to missing data: **Not assessed**	Not assessed.
5. Bias in measurement of the outcome: **Not assessed**	Not assessed.
6. Bias in selection of the reported result: **Not assessed**	Not assessed.
Overall risk of bias: **Critical**	Critical risk of bias due to confounding.

FOR PUBLIC CONSULTATION

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

Supplement 4. Supplementary results

Sensitivity analyses

Sensitivity analysis description / Outcome	No. studies; primary analysis effect estimate (95% CI); τ^2	Description of sensitivity analysis	No. studies; sensitivity analysis effect estimate (95% CI); τ^2
1. Sensitivity analysis. Investigation of the robustness of the meta-analysis estimate to the assumed choice of the imputed correlation coefficient (= 0.5) in uncontrolled before-after studies.			
1a. Gender dysphoria	3; -0.06 (-0.25 to 0.13); $\tau^2 = 0.00$	Imputed correlation coefficient = 0	3; -0.06 (-0.33 to 0.21); $\tau^2 = 0.00$
		Imputed correlation coefficient = 0.3	3; -0.06 (-0.28 to 0.17); $\tau^2 = 0.00$
		Imputed correlation coefficient = 0.7	3; -0.06 (-0.21 to 0.09); $\tau^2 = 0.00$
1b. Body image and satisfaction	3; 0.15 (-0.03 to 0.33); $\tau^2 = 0.00$	Imputed correlation coefficient = 0	3; 0.15 (-0.10 to 0.40); $\tau^2 = 0.00$
		Imputed correlation coefficient = 0.3	3; 0.15 (-0.06 to 0.36); $\tau^2 = 0.00$
		Imputed correlation coefficient = 0.7	3; 0.15 (0.01 to 0.29); $\tau^2 = 0.00$
2. Sensitivity analysis. Investigation of the robustness of the meta-analysis estimate to the exclusion of studies judged to be at a critical risk of bias (RoB)			
2a. Social functioning	4; -0.30 (-0.45 to -0.15); $\tau^2 = 0.00$	Removal of Becker-Hebly 2021 (critical RoB)	3; -0.27 (-0.42 to -0.11); $\tau^2 = 0.00$
2b. Psychological functioning	4; -0.39 (-0.68 to -0.11); $\tau^2 = 0.01$	Removal of Becker-Hebly 2021 (critical RoB)	3; -0.39 (-0.88 to 0.09); $\tau^2 = 0.02$

FOR PUBLIC CONSULTATION

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

Supplement 5. Abbreviations

Abbreviation	Full term / explanation
AFAB	Assigned female at birth
AMAB	Assigned male at birth
CI	Confidence interval
CI_{upper}	Confidence interval upper limit
CI_{lower}	Confidence interval lower limit
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)
GD	Gender dysphoria
GGDC	Gender Guidelines Governance 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
RCT / RT	Randomised controlled trials / randomised trials
ROBINS-I	Risk of bias in non-randomised studies of interventions
SAB	Sex assigned at birth
SD	Standard deviation

FOR PUBLIC CONSULTATION

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

Abbreviation	Full term / explanation
SMD	Standardised mean difference
SR	Systematic review
TGA	Therapeutic Goods Administration
UK	United Kingdom
USA / US	United States of America
WHO	World health organisation
WPATH	World Professional Association for Transgender Health

FOR PUBLIC CONSULTATION

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

References used in the supplement

1. Miller JJ. The Inverse of the Freeman – Tukey Double Arcsine Transformation. The American Statistician. 1978;32(4):138–.
2. Murray FF, John WT. Transformations Related to the Angular and the Square Root. The Annals of Mathematical Statistics. 1950;21(4):607–11.
3. Aloe AM, Thompson CG, Liu Z, Lin L. Estimating Partial Standardized Mean Differences from Regression Models. The Journal of Experimental Education. 2022;90(4):898–915.
4. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Medical Research Methodology. 2014;14(1):135.
5. Higgins JPT, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect [last updated August 2023]. 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 Cochrane, 2024 Available from cochraneorg/handbook2024.
6. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med. 2000;19(22):3127–31.

FOR PUBLIC CONSULTATION

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