Gender-affirming hormone therapy for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis

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

Word Count: Abstract: 244; Total: 3000
Number of Tables: 4
Number of Figures: 1
Number of Appendices: 16

Gender-affirming hormone therapy for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis

Abstract:

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

Objective: In this systematic review and meta-analysis, we assess and summarize the certainty of the evidence about the effects of gender-affirming hormone therapy (GAHT) in individuals experiencing gender dysphoria (GD).

Methods: We searched MEDLINE, Embase, PsychINFO, Social Sciences Abstracts, LGBTQ+ Source, and Sociological Abstracts through September 2023. We included studies comparing GAHT to no GAHT in individuals under 26 years of age experiencing GD. Outcomes of interest included psychological and physical. Pairs of reviewers independently screened articles, abstracted data, and assessed the risk of bias in included studies. We performed meta-analyses and assessed the certainty of the evidence using the GRADE approach.

Results: We included 24 studies. Comparative observational studies (n=9) provided mostly very low certainty evidence regarding gender dysphoria, global function, and depression. One comparative observational study reported that the odds of depression may be lower (OR 0.73 [95% CI 0.61 to 0.88], n (number of studies) =1, low certainty) in individuals who received

GAHT compared to those who did not. Before-after studies (n=13) provided very low certainty evidence about gender dysphoria, global function, depression, and BMD. Case series studies (n=2) provided high certainty evidence that the proportion of individuals with cardiovascular events 7-109 months after receiving GAHT was 0.04 (95% CI 0.03 to 0.05, n = 1, high certainty).

Conclusion: There is considerable uncertainty about the effects of GAHT, and we cannot exclude the possibility of benefit or harm. Methodologically rigorous prospective studies are needed to produce higher certainty evidence.

Key messages:

1. *What is known on this topic:* Previously published evidence syntheses addressing the effects of GAHT in individuals experiencing GD are methodologically limited.
2. *What this study adds:* This publication addresses the effects of GAHT in individuals experiencing GD, while adhering to the highest methodological standards for conducting and reporting a systematic review and meta-analysis, and assessing the risk of bias in each included study and the certainty of the evidence for each outcome of interest.
3. *How this study might affect research, practice, and policy*: The evidence from this systematic review and meta-analysis can be used to inform individuals experiencing GD and considering GAHT, clinicians involved in their care as well as clinical practice guideline developers, policy makers and stakeholders who make decisions about treatment related to gender dysphoria.

Key words: gender-affirming hormone therapy, gender dysphoria, depression, global function, sexual dysfunction, bone mineral density, suicide, cardiovascular event

Introduction

Gender dysphoria (GD) refers to the intense distress caused by feelings of incongruence between one's birth-assigned sex and gender identity. [1] Individuals who experience persistent GD may seek hormonal and surgical interventions to align their physical bodies with their internal or expressed gender and alleviate this distress. [2]

Hormonal treatments for GD in youth include gonadotropin releasing hormone analogues (GnRHa) and gender-affirming hormone therapy (GAHT). The former, GnRHa (puberty blockers), may be administered as early as Tanner Stage 2, [3] followed by GAHT in adolescence to induce and maintain the desired secondary sex characteristics. Hormone therapies include the administration of testosterone for natal females (NF) to create a masculinized appearance, and estrogen in conjunction with GnRHa for natal males (NM) to produce a feminized appearance. Early interventions with puberty blockers followed by GAHT is believed to result in better physical outcomes aligned with the desired gender, [4,5] though some individuals may receive only GAHT.

A high-quality SR is needed to overcome methodological limitations in this field. This SR and meta-analysis aimed to summarize the effects of GAHT in individuals with GD under the age of 26.

Methods

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

Eligibility criteria

For eligibility criteria, see Table 1.

Table 1. Eligibility criteria.

Types of studies	We included randomized controlled trials, comparative observational studies, and before-after studies addressing the intervention and comparison of interest. We also included case series addressing the intervention of interest in special instances. We did not find any RCTs and included all eligible comparative observational and before-after studies. As for case series, if an outcome of interest was not reported in the eligible comparative observational or before-after studies, we included

	all eligible case series studies addressing that outcome. We included studies published in full, and in English language.
Population	We included individuals under 26 years, who were diagnosed, experienced, self-identified, or were identified by a parent as having GD, gender identity disorder or gender incongruence, or who identified as transgender or non-binary. To be as inclusive as possible, we included all studies where the mean age of participants was below 26 years. We decided to include individuals below 26 years of age because the definition of youth, the target population of this review, is commonly defined as extending into the mid-twenties. [6,7]
Intervention	We included studies assessing the effects of GAHT. We defined GAHT as stated by the authors or as the use of feminizing hormones in an individual assigned male at birth or as the use of masculinizing hormones in an individual assigned female at birth.
Comparator	The comparator of interest was no GAHT (e.g., psychological therapy, no treatment). In case series studies, a comparator group was not necessary.
Outcomes	We included studies reporting on any of the following outcomes if follow up was short term (≤6 months) or long-term (≥1 year): gender dysphoria, completed suicides, global function, depression, sexual dysfunction from physiological perspective (i.e., lack of erection, dyspareunia, problems related to dry and degenerated mucosal tissue, anorgasmia), bone mineral density (BMD), and cardiovascular events.

Information sources

With the assistance of an information specialist (RC), we searched MEDLINE, Embase, PsycINFO, Social Sciences Abstracts, Contemporary Women's Issues, LGBTQ+ Source, Sociological Abstracts, Studies on Women, Gender Abstracts, and Google Scholar from inception to September 2023. This search was part of an umbrella search for a related SR. [8] All search strategies are in Appendix 2.

Study selection

Two reviewers (SI, YR), using Covidence software (https://www.covidence.org/) and following training and calibration exercises, independently screened titles and abstracts, and full texts of potentially eligible studies. A third reviewer (AM) resolved conflicts. The study selection for this SR was completed in tandem with another related SR at the abstract and full-text stages. [8]

Data collection

For data collection, see Appendix 3.

Risk of bias in individual studies

We assessed the risk of bias using a modified version of the Cochrane risk of bias tool for non-randomized studies of interventions (ROBINS-I) for each study design. For details, see Appendix 4 and 5.

Data synthesis

Although study authors used various observational study designs, we classified studies according to how the data were analyzed for this review. See Appendix 6.

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

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

We conducted meta-analysis using a random-effects model when appropriate, as determined by subject area experts (CKM, SM), for studies addressing the same outcome and with no clinical heterogeneity between them (i.e., study design, population, intervention/comparator, outcome

definition). When studies reported outcomes using different scales, we calculated the standardized mean change for before-after studies. If we could not perform a meta-analysis, we summarized the evidence across studies. We used the *meta* and *metafor* packages in R Studio Version 4.2 for analyses.

Certainty of the evidence

We assessed the certainty of the evidence following grading of recommendations assessment, development, and evaluation (GRADE) approach. [9] For details, see Appendix 7. We assessed the certainty in the causal effects of GAHT on the outcomes of interest rather than the association between GAHT as an exposure. We followed GRADE guidance and principles to address questions about interventions using observational studies. This process involves clarifying the question (target of certainty), defining the intent of the question (causality), and assessing the certainty of the evidence under those parameters.

When assessing risk of bias for each outcome, we rated down the certainty of evidence from observational studies by up to three levels due to prognostic imbalance. In case series, outcomes requiring a comparison group (e.g., GD, completed suicides, global function, depression, sexual dysfunction, BMD) were rated down three levels due to the absence of such a group. However, outcomes not requiring a comparison group (e.g., cardiovascular events linked to gender-affirming hormones) were not rated down, as these events were specific to intervention recipients.

To minimize value judgments, we used a null effect threshold (1 for relative measures, and 0 for absolute measures, mean differences, or mean changes) to rate the certainty of any benefit or harm (of any magnitude) from receiving GAHT over not receiving GAHT. We did not define a minimally important difference to determine whether an effect was clinically meaningful or important.

Subgroup and sensitivity analyses

For subgroup and sensitivity analyses, see Appendix 8.

Management of conflicts of interest

For management of conflicts of interest, see Appendix 9. Other SRs under the described agreement include SRs about the effects of social gender transition, mastectomy, [10] chest binding and genital tucking, and puberty blockers (all submitted for publication).

Results

After screening 6,736 titles and abstracts for this SR and another related , [8] we included 24 studies. Figure 1 shows the study search and selection process. We present reasons for exclusion at the full-text screening stage (n=311) with references in Appendix 10.

Characteristics of included studies

Of 24 included studies, 9 were comparative observational, [11-19] 13 were before-after, [20-32] and 2 were case series [33,34] (Figure 1). Thirteen studies included NMs and NFs, and 11 included NFs only.

The mean (SD) age of participants at the time of GAHT ranged from 15.1 (1.8) to 25.1 (4.8). We present characteristics of included studies in Appendix 11. Appendix 12 describes outcome measurement instruments used in the studies and their interpretability.

Risk of bias in included studies

Across comparative observational studies, the domains commonly judged as serious or critical risk of bias were confounding, missing data, and deviation from intended intervention (i.e., administration of co-interventions). Before-after studies were at serious or critical risk of bias due to missing data and deviation from intended intervention (i.e., administration of co-interventions). In addition to lacking a comparison group, case series were at critical risk of bias due to measurement of the outcome (Appendix 13).

Effects of interventions

We describe the effects of the interventions for each study design. Tables 1-3 provide summary of findings tables, and appendix 14 displays forest plots of meta-analysis. If sex-specific data were available, we included separate data points for NMs and NFs in each meta-analysis

(Appendix 14). When studies reported data for both groups and no important heterogeneity was found, we presented a single combined effect estimate.

1. Comparative observational studies

See Table 2 for summary of findings table.

Gender dysphoria (GD): Current GD, using the Gender Distress Scale ranging from 1 to 5, may be lower (MD (mean difference) 0.4 lower [95% CI 0.24 lower to 0.16 higher], number of studies (n) = 1, very low certainty) in NMs and NFs who received GAHT compared to those who did not; however, we are very uncertain about the causal effect of the intervention on GD. [15]

Global function: A meta-analysis suggests that global function, measured within the last 12 to 24 months, may be higher (standardized mean difference (SMD) 0.87 higher, [95% CI 0.25 lower to 2 higher], n = 2, very low certainty) in NMs and NFs who received GAHT compared to those who did not; however, we are very uncertain about the causal effect of the intervention on global function. [35,36]

Depression: Eight studies reported this outcome using seven different measurement instruments. Due to variability in instruments, time points, and reporting, we could not include all studies in a single meta-analysis.

A meta-analysis suggests that depression, measured within the last 12 months, may be lower (SMD 0.3, [95% CI 0.85 lower to 0.25 higher], n = 2, very low certainty) in NMs and NFs who received GAHT compared to those who did not; however, we are very uncertain about the causal effect of the intervention on depression. [16,36] See Appendices 15 and 16 for low to very low certainty evidence about depression from studies not pooled with this evidence.

Table 2. Gender affirming hormone therapy vs no gender affirming hormone therapy: evidence from comparative observational studies.

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with no gender affirming hormone therapy	Risk with gender affirming hormone therapy				
Gender Dysphoria, current (no follow-up) assessed with: participant reported Gender Distress Scale, higher scores indicate higher gender distress Scale from: 1 to 5	The mean gender Dysphoria, current (no follow-up) was **4.17**	MD **0.4 lower** (0.24 lower to 0.16 higher)	-	146 (1 non-randomised study)[1]	⨁◯◯◯ Very low [a, b]	The evidence is very uncertain about the effect of gender affirming hormone therapy on gender dysphoria (no follow-up) in natal males and natal females.
Global Function, Long Term Follow-Up assessed with: participant reported, various scales [Symptom Checklist-90 Revised (SCL-90-R) Global Severity Index, The Children's Global Assessment Scale (CGAS)], higher scores indicate better global function follow-up: range 12 months to 24 months [c]	-	SMD **0.87 SD higher** (0.25 lower to 2 higher)	-	125 (2 non-randomised studies)[2,3]	⨁◯◯◯ Very low [d, e, f]	The evidence is very uncertain about the effect of gender affirming hormone therapy on global function at long term follow-up in natal males and natal females.

Depression, Long Term Follow-Up assessed with: participant reported, various scales [Symptom Checklist-90 Revised (SCL-90-R) Depression Domain, Children's Depression Inventory (CDI)], higher scores represent worse depression follow-up: mean 12 months [c]	-	SMD **0.3 SD lower** (0.85 lower to 0.25 higher)	-	154 (2 non-randomised studies)[2,4]	⨁◯◯◯ Very low [f, g, h]	The evidence is very uncertain about the effect of gender affirming hormone therapy on depression at long term follow-up in natal males and natal females.
Other Outcomes - not measured [i]	-	-	-	-	-	-

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

CI: confidence interval; **MD:** mean difference; **OR:** odds ratio; **SMD:** standardised mean difference

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

Explanations

a. Rated down three levels due to critical risk of bias because of lack of adjustment for important confounders (i.e., psychiatric interventions, mental health comorbidities, socioeconomic status, and family support) and missing data (i.e., 41.71% provided outcome data).
b. Rated down one level for imprecision as the confidence intervals cross the threshold of no effect (i.e., MD=0), suggesting both a possibility of a benefit or harm in the outcome.
c. Long Term Follow-Up: outcome measured at ≥ 12 months follow-up.
d. Rated down three levels due to critical risk of bias because of lack of adjustment for important confounders in the two included studies (i.e., psychiatric interventions, mental health comorbidities, socioeconomic status, and family support) and missing data in one of the two included studies (i.e., 37% provided outcome data).
e. Statistically, there was considerable heterogeneity with I2=88% and p<0.01. However, we did not rate down for inconsistency as this heterogeneity could be explained by the fact that one of the two included studies measured the outcome only in natal female participants, while the other study measured the outcome in natal female and male participants.
f. Rated down one level for imprecision as the confidence intervals cross the threshold of no effect (i.e., SMD=0), suggesting both a possibility of a benefit or harm in the outcome.
g. Rated down three levels due to critical risk of bias because of lack of adjustment for important confounders in the two included studies (i.e., psychiatric interventions, mental health comorbidities, socioeconomic status, and family support) and serious risk of bias due to deviation of intended intervention in one of the included studies (28.26% of the participants in the no gender-affirming hormone therapy group were receiving puberty blockers or spironolactone as mono-therapy).
h. Statistically, there was moderate heterogeneity with I2=63% and p=0.03. However, we did not rate down for inconsistency as this can be explained because one of the two included studies measured the outcome only in natal female participants, while the other study measured in natal female and male participants.
i. Outcomes not measured: death by suicide, sexual dysfunction from a physiological perspective (i.e., lack of erection, dyspareunia, problems related to dry and degenerated mucosal tissue, anorgasmia), bone density, cardiovascular events.

References

1.AMV, Fontanari, F, Vilanova, MA, Schneider, I, Chinazzo, BM, Soll, K, Schwarz, MIR, Lobato, A, Brandelli, Costa. Gender affirmation is associated with transgender and gender nonbinary youth mental health improvement. LGBT Health; 2020.
2.H, Filipov, Y, Kavla, S, Şahin, ME, Gökler, Ş, Turan. The effects of gender-affirming hormone therapy on body satisfaction, self-esteem, quality of life, and psychopathology in people with female-to-male gender dysphoria. Transgender Health; 2023.
3.I, Becker-Hebly, S, Fahrenkrug, F, Campion, H, Richter-Appelt, M, Schulte-Markwort, C, Barkmann. Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: a descriptive study from the Hamburg Gender Identity Service European Child and Adolescent Psychiatry; 2021.
4.C, Grannis, WI, Mattson, SF, Leibowitz, L, Nahata, D, Chen, JF, Strang, H, Thobe, JA, Indyk, EE, Nelson. Expanding upon the relationship between gender-affirming hormone therapy, neural connectivity, mental health, and body image dissatisfaction. Psychoneuroendocrinology; 2023.

2. Before-after studies

See Table 3 for summary of findings table.

Gender dysphoria (GD): Meta-analysis suggested that GD, measured within the last 6 months with the Gender Preoccupation and Stability Questionnaire ranging from 14 to 70, was lower (SMD 0.26 lower [95% CI 1.64 lower to 1.13 higher], n = 2, very low certainty) in NFs after receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on GD. [21,23]

Global function: Three studies reported global function using 3 different measures at 2 different timepoints.

Global function, measured within the last 6 months, may be higher (SMD 0.25 higher [95% CI 0.09 higher to 0.4 higher], n = 2, very low certainty) in NFs after receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on global function. [23,27] See Appendices 15 and 16 for very low certainty evidence about global function from studies not pooled with this evidence.

Depression: Four studies reported this outcome using 4 different scales. Due to variability in measurement instruments, timepoints, and reporting, we could not include all studies in a single meta-analysis.

A meta-analysis suggested that depression, measured within 18 to 24 months, may be lower (SMD 0.41 lower [95% CI 0.65 lower to 0.17 lower], n = 2, very low certainty) in NMs and NFs after receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on depression. [20,22] See Appendices 15 and 16 for very low certainty evidence about depression from studies not pooled with this evidence.

Sexual dysfunction: A study reported a linear regression analysis with no statistically significant change in sexual dysfunction (i.e., vagina dryness or itch) reported by the NFs after 6 months of receiving GAHT (b = -0.01, 95% CI -0.09, 0.8) and after 12 months of receiving GAHT (b= 0.053, 95% CI: -0.03, 0.13) compared to before the intervention. This evidence was rated as low certainty; therefore, we are very uncertain about the causal effect of the intervention on sexual dysfunction. [28]

Bone mineral density (BMD): Six studies reported lumbar spine BMD, 3 studies reported femoral neck BMD, and 3 studies reported hip BMD using z-scores and g/cm^2.

Lumbar spine BMD, measured within the last 12 to 36 months with g/cm^2 may be higher (0.01 higher [95% CI 0 higher to 0.01 higher], n=2, very low certainty) in NFs receiving GAHT

compared to before, although we are very uncertain about the causal effect of the intervention on lumbar spine BMD. [29,32]

Femoral neck BMD, measured within the last 12 months assessed with the DXA, z-scores ranging from -3 to 3, may not change (MC 0 [95% CI 0.01 lower to 0], n=1, very low certainty) in NFs after receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on femoral neck BMD.[32]

Hip BMD, measured within the last 12 to 36 months with g/cm^2 was higher (0.01 higher [95% CI 0.01 higher to 0.01 higher], n=1, very low certainty) in NFs receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on hip BMD. [32]

See Appendices 15 and 16 for very low certainty evidence about BMD from studies not pooled with this evidence.

Table 3. Gender affirming hormone therapy vs no gender affirming hormone therapy: evidence from before-after studies.

Outcomes	**Anticipated absolute effects*** (95% CI)		№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk with no gender affirming hormone therapy	**Risk with gender affirming hormone therapy**			

Gender Dysphoria, Short Term Follow-up assessed with: participant reported, Gender Preoccupation and Stability Questionnaire, higher scores indicate higher levels of gender dysphoria Scale from: 14 to 70 follow-up: mean 6 months[a]	-	standardized mean change **0.26 lower** (1.64 lower to 1.13 higher)	36 (1 non-randomised study)[1,2]	⨁◯◯◯ Very low[b,c]	The evidence is very uncertain about the effect of gender affirming hormone therapy on gender dysphoria at short term follow-up in natal females.
Global Function, Short Term Follow-up assessed with: participant reported, various scales [RAND Short Form-36 (SF-36) Health Survey, Symptom Checklist-90 Revised (SCL-90-R) Global Severity Index], higher scores indicate better global function Scale from: 0 to 100 follow-up: mean 6 months[a]	-	standardized mean change **0.25 higher** (0.09 higher to 0.4 higher)	73 (2 non-randomised studies)[1,3]	⨁◯◯◯ Very low[e,d,e,f]	The evidence is very uncertain about the effect of gender affirming hormone therapy on global function at short term follow-up in natal females.
Depression, Long Term Follow-up assessed with: participant reported, various scales [Beck Depression Inventory, Hospital Anxiety and Depression Scale (HADS)], higher scores indicate worse depression follow-up: range 18 months to 24 months[g]	-	standardized mean change **0.41 lower** (0.65 lower to 0.17 lower)	389 (2 non-randomised studies)[4,5]	⨁◯◯◯ Very low[h,i]	The evidence is very uncertain about the effect of gender affirming hormone therapy on depression at long term follow-up in natal males and females.
Sexual Dysfunction (i.e., Vaginal Dryness or Itch), Long Term Follow-Up assessed with: participant report of symptoms follow-up: mean 12 months[g]	In 193 participants, a linear regression analysis showed that there was no change from baseline in symptoms of vaginal dryness or itch after receiving GAHT (b= 0.053, 95% CI: -0.03, 0.13). [j,k]		193 (1 non-randomised study)[6]	⨁◯◯◯ Very low[l]	The evidence is very uncertain about the effect of gender affirming hormone therapy on sexual dysfunction (i.e., vaginal dryness or itch) at long term follow-up in natal females.
Sexual Dysfunction (i.e., Vaginal Dryness or Itch), Short Term Follow-Up assessed with: participant report of symptoms follow-up: mean 6 months[a]	In 193 participants (i.e., natal females), a linear regression analysis showed that there was no change from baseline in symptoms of vaginal dryness or itch after receiving GAHT (b= -0.01, 95% CI: -0.09, 0.8). [j,k]		193 (1 non-randomised study)[6]	⨁◯◯◯ Very low[l]	The evidence is very uncertain about the effect of gender affirming hormone therapy on sexual dysfunction (i.e., vaginal dryness or itch) at short term follow-up in natal females.
Bone Mineral Density - Femoral Neck, Long Term Follow-up assessed with: Dual-energy x-ray absorptiomety (DXA), z-scores Scale from: -3 to 3 follow-up: mean 12 months[g]	The mean bone Mineral Density - Femoral Neck, Long Term Follow-up was **0.84**	mean change **0** (0.01 lower to 0)	199 (1 non-randomised study)[7]	⨁◯◯◯ Very low[m]	The evidence is very uncertain about the effect of gender affirming hormone therapy on bone mineral density - femoral neck at long term follow-up in natal females.

Bone Mineral Density - Hip, Long Term Follow-up assessed with: Dual-energy x-ray absorption (DXA), g/cm2 follow-up: range 12 months to 36 months[g]	The mean bone Mineral Density - Hip, Long Term Follow-up was **0.95**	mean change **0.01 higher** (0.01 higher to 0.01 higher)	199 (1 non-randomised study)[7]	⨁◯◯◯ Very low[m]	Gender affirming hormone therapy may increase bone mineral density - hip, at long term follow-up slightly in natal females.
Bone Mineral Density - Lumbar Spine, Long Term Follow-up assessed with: Dual-energy x-ray absorption (DXA), g/cm2 follow-up: range 12 months to 36 months[g]	The mean bone Mineral Density - Lumbar Spine, Long Term Follow-up was **1.04**	mean change **0.01 higher** (0 to 0.01 higher)	234 (2 non-randomised studies)[7,8]	⨁◯◯◯ Very low[m]	Gender affirming hormone therapy may increase bone mineral density - lumbar spine, at long term follow-up slightly in natal females.
Other Outcomes - not measured[n]	-	-	-	-	-

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

CI: confidence interval; **MD:** mean difference; **OR:** odds ratio; **SMD:** standardised mean difference

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

Explanations

a. Short Term Follow-Up: outcome measured at ≤ 6 months follow-up.
b. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design and critical risk of bias due to missing data (i.e., 46.75% provided outcome data).
c. Rated down one level for imprecision because the optimal information size (OIS=200) was not met. Low sample size importantly increases the risk of random error.
d. Rated down two levels due to risk of bias stemming from prognostic imbalance associated with the observational study design and critical risk of bias due to missing data in one of the two included studies (i.e., 46.75% provided outcome data).
e. Statistically, there was considerable heterogeneity with I2=94% and p<0.01. However, we did not rate down for inconsistency as the overall effect estimate was not importantly affected by the studies contributing to statistical heterogeneity.
f. Rated down one level for indirectness because one of the two included studies reports the outcome only for natal females.
g. Long Term Follow-Up: outcome measured at ≥ 12 months follow-up.
h. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design, as well as critical and serious risk of bias due to missing data in the two included studies (i.e., 20% and 69% of participants, respectively, provided outcome data).
i. Statistically, there was considerable heterogeneity with I2=100% and p<0.01. However, we did not rate down for inconsistency as the overall effect estimate was not importantly affected by the studies contributing to statistical heterogeneity.
j. GAHT: gender affirming hormone therapy.
k. In the linear mixed model, time was as added categorical variable to detect changes in symptom scores between 0-3 months, 0-6 months, and 0-12 months of GAHT. Differences in changes in symptom scores between different administration forms were corrected for baseline differences to avoid regression to the mean. An increase or decrease in symptom scores of 0.2 was considered clinically relevant.
l. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design and critical risk of bias due to concerns with measurement of the outcome (i.e., subjective and self-reported outcome).
m. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design and critical risk of bias due to missing data (i.e., 48% of participants provided outcome data).
n. Other outcomes: gender dysphoria, sexual dysfunction from physiological perspective (i.e., lack of erection, dyspareunia, anorgasmia), cardiovascular events.

References

1.L, Foster, Skewis, I, Bretherton, SY, Leemaqz, JD, Zajac, AS, Cheung. Short-term effects of gender-affirming hormone therapy on dysphoria and quality of life in transgender individuals: a prospective controlled study. Frontiers in Endocrinology; 2021.
2.M, Ascha, DC, Sasson, R, Sood, JW, Cornelius, JM, Schauer, A, Runge, AL, Muldoon, N, Gangopadhyay, L, Simons, D, Chen, JF, Corcoran, SW, Jordan. Top surgery and chest dysphoria among transmasculine and nonbinary adolescents and young adults. JAMA pediatrics; 2022.
3.Ş, Turan, C, Aksoy,Poyraz, NG, Usta,Sağlam, ÖF, Demirel, Ö, Haliloğlu, P, Kadıoğlu, A, Duran. Alterations in body uneasiness, eating attitudes, and psychopathology before and after cross-sex hormonal treatment in patients with female-to-male gender dysphoria. Archives of Sexual Behavior; 2018.
4.D, Chen, J, Berona, YM, Chan, D, Ehrensaft, R, Garofalo, MA, Hidalgo, SM, Rosenthal, AC, Tishelman, J, Olson-Kennedy. Psychosocial functioning in transgender youth after 2 years of hormones. The New England Journal of Medicine; 2023.
5.Z, Aldridge, S, Patel, B, Guo, E, Nixon, W, Pierre, Bouman, GL, Witcomb, J, Arcelus. Long-term effect of gender-affirming hormone treatment on depression and anxiety symptoms in transgender people: A prospective cohort study. Andrology; 2021.
6.D, van,Dijk, MJHJ, Dekker, EB, Conemans, CM, Wiepjes, EGM, de,Goeij, KA, Overbeek, AD, Fisher, M, den,Heijer, G, T'Sjoen. Explorative prospective evaluation of short-term subjective effects of hormonal treatment in trans people-results from the European network for the investigation of gender incongruence. The Journal of Sexual Medicine; 2019.
7.CM, Wiepjes, MC, Vlot, M, Klaver, NM, Nota, CJ, de,Blok, RT, de,Jongh, P, Lips, AC, Heijboer, AD, Fisher, T, Schreiner, G, T'Sjoen, M, den,Heijer. Bone mineral density increases in trans persons after 1 year of hormonal treatment: a multicenter prospective observational study. Journal of Bone and Mineral Research : The Official Journal of the American; 2017.
8.P, van,Kesteren, P, Lips, W, Deville, C, Popp-Snijders, H, Asscheman, J, Megens, L, Gooren. The effect of one-year cross-sex hormonal treatment on bone metabolism and serum insulin-like growth factor-1 in transsexuals. The Journal of Clinical Endocrinology and Metabolism; 1996.

3. Case series

See Table 4 for summary of findings table. One of the before-after studies reported data about death by suicide only after the intervention and we classified it as case series for that outcome. [22]

Death by suicide: As retrieved from medical records, death by suicide within 24 months of receiving GAHT occurred in 2 of 315 NMs and NFs (0.6%); proportion 0.006 (95% CI 0.001 to 0.018, n = 1, very low certainty). We are very uncertain about the effects of GAHT on death by suicide. [22]

Cardiovascular events: As retrieved from medical records, cardiovascular events within 7 to 109 months of receiving GAHT occurred in 151 of 3875 NFs (3.9%); proportion 0.04 (95% CI 0.03 to 0.05, n = 1, high certainty). [33]

Cardiovascular events: As retrieved from medical records, cardiovascular events within 26 months of receiving GAHT occurred in 3 of 1893 NFs (0.2%); proportion 0.00 (95% CI 0.00 to 0.01, n = 1, moderate certainty). [34]

Table 4. Gender affirming hormone therapy vs no gender affirming hormone therapy: evidence from case series.

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with no gender affirming hormone therapy	Risk with gender affirming hormone therapy				
Death by Suicide, Long Term Follow-up assessed with: medical records follow-up: mean 24 months [a]	No comparison group available	**6 per 1,000** (1 to 18)	**proportion 0.006** (0.001 to 0.018)	315 (1 non-randomised study)[1]	⨁◯◯◯ Very low [b]	The evidence is very uncertain about the effect of gender affirming hormone therapy on death by suicide at long term follow-up in natal males and females.
Cardiovascular Events, Long Term Follow-Up assessed with: medical records, number of events follow-up: range 7 months to 109 months [a, c]	No comparison group available	**40 per 1,000** (30 to 50)	**proportion 0.04** (0.03 to 0.05)	3875 (1 non-randomised study)[2]	⨁⨁⨁⨁ High[f]	The proportion of natal females experiencing cardiovascular events at long term follow-up is 40 per 1,000.
Cardiovascular Events, Long Term Follow-Up assessed with: medical records, number of participants with an event follow-up: mean 26 months [a, d]	No comparison group available	**0 per 1,000** (0 to 10)	**proportion 0.00** (0.00 to 0.01)	1893 (1 non-randomised study)[3]	⨁⨁⨁◯ Moderate [ef]	The proportion of natal females experiencing cardiovascular events at long term follow-up is 1 per 1,000.
Other Outcomes - not measured [g]	-	-	-	-	-	-

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

CI: confidence interval

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

Explanations

a. Long Term Follow-Up: outcome measured at ≥ 12 months follow-up.
b. Rated down three levels for risk of bias due to lack of a comparison group.
c. Cardiovascular events include: stroke, myocardial infarction, and venous thromboembolism.
d. Cardiovascular events include: thromboembolism.
e. Rated down one level for indirectness because this study included natal males only.
f. We did not rate down for risk of bias because this outcome does not need a comparison group, as the study participants can only experience this outcome if they have received the intervention.
g. Other outcomes not measured: gender dysphoria, global function, depression, sexual dysfunction from physiological perspective (i.e., lack of erection, dyspareunia, problems related to dry and degenerated mucosal tissue, anorgasmia), bone mineral density.

References

1.D, Chen, J, Berona, YM, Chan, D, Ehrensaft, R, Garofalo, MA, Hidalgo, SM, Rosenthal, AC, Tishelman, J, Olson-Kennedy. Psychosocial functioning in transgender youth after 2 years of hormones. The New England Journal of Medicine; 2023.
2.NM, Nota, CM, Wiepjes, CJM, de,Blok, LJG, Gooren, BPC, Kreukels, M, den,Heijer. Occurrence of acute cardiovascular events in transgender individuals receiving hormone therapy. Circulation; 2019.
3.M, Pyra, I, Casimiro, L, Rusie, N, Ross, C, Blum, K, Keglovitz, Baker, A, Baker, J, Schneider. An observational study of hypertension and thromboembolism among transgender patients using gender-affirming hormone therapy. Transgender Health; 2020.

Discussion

This SR and meta-analysis synthesized the available evidence regarding the effects of GAHT in young individuals with GD. Comparative observational studies provided mostly very low certainty evidence for GD, global function, and depression. One study provided low certainty evidence that depression may be lower in NMs and NFs who received GAHT compared to those who did not. Before-after studies provided very low certainty evidence. Case series provided very low certainty evidence on death by suicide and high to moderate certainty evidence for cardiovascular events.

Although some may view our modification of the ROBINS-I tool as a limitation, we strongly believe it produced similar conclusions than if we had used the original tool or alternatives, such as the Newcastle-Ottawa scale. [37] Given the widespread methodological limitations in this field, any risk of bias tool would yield similar conclusions. Comparative observational and before-after studies were at serious or critical risk of bias due to missing data and deviation from intended intervention (i.e., administration of co-interventions). Case series, which lack a comparison group, were at critical risk of bias due to measurement of the outcome. These studies should only be used to generate hypothesis for more rigorous study designs, such as prospective cohorts.

The target question of this SR – and of the decision-makers considering these interventions– is: what are the effects of GAHT? In the absence of randomized controlled trials or comparative observational studies, case series and before-after studies provide the best available evidence to answer this question. While these study designs answer single-group questions (e.g., what is the functional status among people who received GAHT), they are limited in assessing intervention effects (e.g., whether functional status is better in people who received GAHT than those who did not). We accounted for these limitations, and assessed the certainty of the available evidence following current methodological standards. [38]

We rated down the certainty of the evidence mostly because of risk of bias and imprecision, often resulted from an insufficient sample size, for most outcomes and study designs. We did not find evidence about sexual dysfunction in NMs.

The overarching theme from this and other SRs on GAHT is the lack of high-quality evidence for individuals with GD. Unlike this SR, other reviews did not assess the certainty of evidence for each outcome.

Taylor et al focused on individuals 18 years and below, rating most studies as low to moderate quality using the Newcastle-Ottawa Scale. They found limited evidence on GD, body satisfaction, psychological and cognitive outcomes, and infertility. [39] Doyle et al reported on psychosocial functioning changes after GAHT among transgender individuals of all ages. They concluded that risk of bias, assessed with the Newcastle-Ottawa Scale, varied among studies. Small sample sizes and undadjusted confounders limited the ability to draw causal inferences. [40]

Van Leerdam et al concluded that GAHT may reduce GD, body dissatisfaction, and uneasiness, subsequently improving psychological well-being and quality of life in transgender individuals of all ages. [41] They rated the evidence as low to moderate in quality, based on longitudinal cohort and cross-sectional studies, without clarifying their rating methods. Chew et al suggested that GAHT helps adolescents achieve intended physical effects, with limited evidence on its psychosocial and cognitive impact. [42] Further, a SR by Connelly et al concluded that current data are insufficient to determine GAHT's impact on blood pressure in transgender individuals. [43]

Across all these SRs, the findings highlight methodological limitations, low-quality evidence, and significant gaps.

The evidence about the effects of GAHT in individuals under the age of 26 experiencing GD is predominantly very low certainty, with lack of moderate and high certainty evidence about the effects of this intervention. This information is crucial for patients, caregivers, clinicians, guideline developers, and policymakers involved in treatment decisions. Beyond evidence certainty, decision-making should consider other factors, including the magnitude and consequences of potential benefits and harms, patients' and caregivers' values and preferences, resource use, feasibility, acceptability, and equity.[44] Guideline developers and policy makers must transparently state which and whose values they prioritize when developing treatment recommendations and policies.

Strengths and limitations of the review process

This SR and meta-analysis has multiple strengths. We rigorously followed the highest methodological standards, we assessed the risk of bias for each study using the ROBINS-I tool, and evaluated the certainty of the evidence for each outcome using the GRADE approach. A limitation of our review is the inclusion of only English-language studies, though we do not expect this to fundamentally alter our conclusions. Due to feasibility considerations, we prioritized specific outcomes and could not address others that may be important to readers, such as regret, anxiety, pelvic pain, or cancers (e.g., breast, gynecological, prostate, and colon cancer).

Conclusion

The best available evidence reporting on the effects of GAHT in individuals experiencing GD ranged from moderate to high certainty for cardiovascular events, and low to very low certainty

for the outcomes of GD, global function, depression, sexual dysfunction, BMD, and death by suicide. We did not find evidence on NM sexual dysfunction. The evidence found in this SR and meta-analysis does not exclude the possibility of benefit or harm upon receipt of GAHT. Prospective studies yielding higher certainty evidence are needed to understand the short- and long-term effects of GAHT.

Contributorship Statement:

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

Funding and Disclosures:

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

1. Funding:

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

2. Financial conflicts of interest (manuscript authors):

Direct financial conflicts of interest:

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

Indirect financial conflicts of interest:

Manuscript authors do not have indirect financial conflicts of interest.

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

Direct financial conflicts of interest:

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

Indirect financial conflicts of interest:

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

3. Other disclosures (manuscript authors):

Expressed opinions:

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

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

Affiliations:

Dr. William Malone is a board member of SEGM.

Expressed opinions:

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

Ethics Approval:

This study does not involve human or animal subjects.

Data Sharing:

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

AI Use Statement:

AI technology was not used in this project.

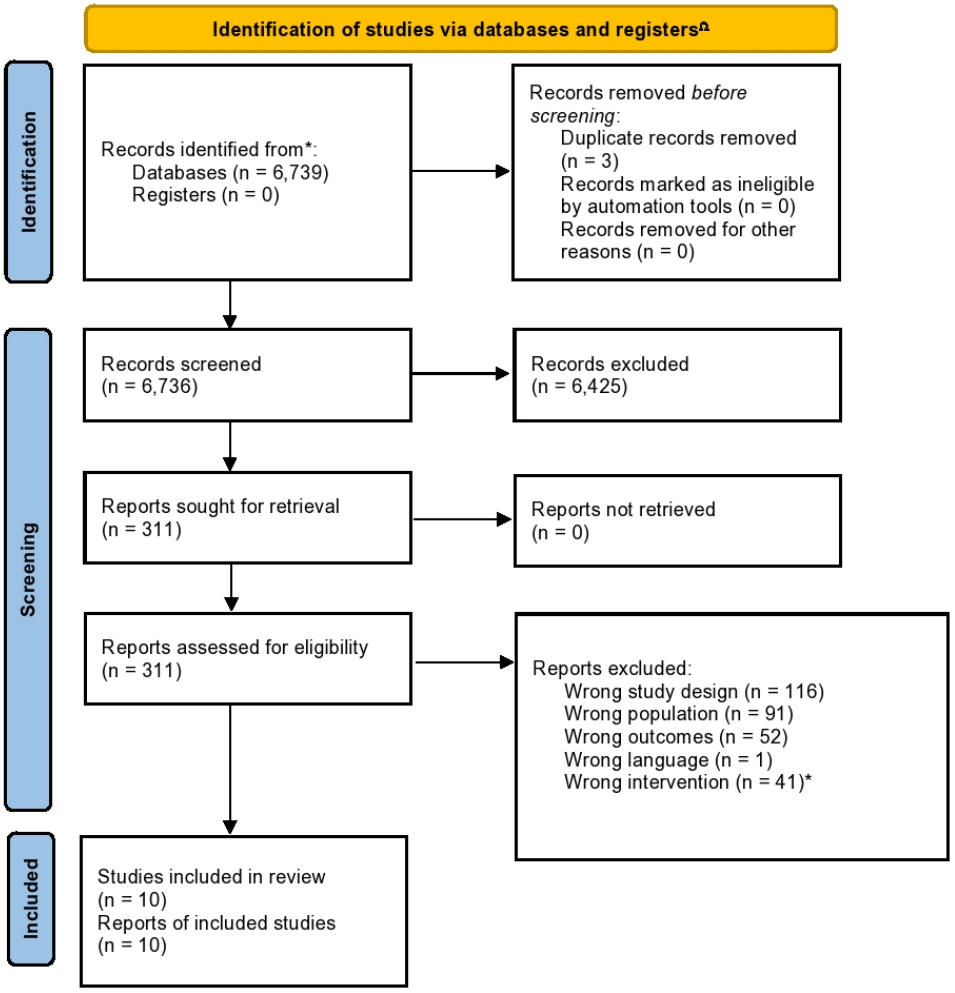
References:

1. American Psychiatric Association. The Diagnostic and Statistical Manual of Mental Disorders (5th ed DSM-5). Arlington, VA: American Psychiatric Association; 2013.
2. Trans Care BC. 2024. https://www.transcarebc.ca/.
3. Hembree WC. Guidelines for pubertal suspension and gender reassignment for transgender adolescents. *Child Adolesc Psychiatr Clin N Am* 2011; **20**(4): 725-32.
4. Biggs M. The Dutch Protocol for Juvenile Transsexuals: Origins and Evidence. *Journal of Sex & Marital Therapy* 2023; **49**(4): 348-68.
5. Cohen-Kettenis PT, Pfäfflin F. The DSM diagnostic criteria for gender identity disorder in adolescents and adults. *Arch Sex Behav* 2010; **39**(2): 499-513.
6. Mercy JA, Vivolo-Kantor AM. The Center for Disease Control and Prevention's (CDC) Youth Violence Prevention Centers: Paving the Way to Prevention. *J Prim Prev* 2016; **37**(2): 209-14.
7. United Nations. Youth. https://www.un.org/en/global-issues/youth.
8. Miroshnychenko A, Roldan Y, Ibrahim S, et al. Puberty blockers for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis. *Submitted for publication* 2024.
9. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; **336**(7650): 924-6.
10. Miroshnychenko A, Roldan YM, Ibrahim S, et al. "Mastectomy for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis". *Plast Reconstr Surg* 2024.
11. Achille C, Taggart T, Eaton NR, et al. Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: preliminary results. *International Journal of Pediatric Endocrinology* 2020; **2020**: 8.
12. Becker-Hebly I, Fahrenkrug S, Campion F, Richter-Appelt H, Schulte-Markwort M, Barkmann C. Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: a descriptive study from the Hamburg Gender Identity Service. *European Child & Adolescent Psychiatry* 2021; **30**(11): 1755-67.
13. 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. *Transgender Health* 2020; **5**(3): 196-200.
14. Filipov H, Kavla Y, Sahin S, Gokler ME, Turan S. The Effects of Gender-Affirming Hormone Therapy on Body Satisfaction, Self-Esteem, Quality of Life, and Psychopathology in People with Female-to-Male Gender Dysphoria. *Transgender Health* 2023; **8**(2): 168-74.
15. Fontanari AMV, Vilanova F, Schneider MA, et al. Gender Affirmation Is Associated with Transgender and Gender Nonbinary Youth Mental Health Improvement. *Lgbt Health* 2020; **7**(5): 237-47.

16. Grannis C, Mattson WI, Leibowitz SF, et al. Expanding upon the relationship between gender-affirming hormone therapy, neural connectivity, mental health, and body image dissatisfaction. *Psychoneuroendocrinology* 2023; **156**: 106319.
17. Green AE, DeChants JP, Price MN, Davis CK. Association of Gender-Affirming Hormone Therapy With Depression, Thoughts of Suicide, and Attempted Suicide Among Transgender and Nonbinary Youth. *Journal of Adolescent Health* 2022; **70**(4): 643-9.
18. Olsavsky AL, Grannis C, Bricker J, et al. Associations Among Gender-Affirming Hormonal Interventions, Social Support, and Transgender Adolescents' Mental Health. *Journal of Adolescent Health* 2023; **72**(6): 860-8.
19. 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 Network Open* 2022; **5**(2): e220978.
20. Aldridge Z, Patel S, Guo B, et al. Long-term effect of gender-affirming hormone treatment on depression and anxiety symptoms in transgender people: A prospective cohort study. *Andrology* 2021; **9**(6): 1808-16.
21. Ascha M, Sasson DC, Sood R, et al. Top Surgery and Chest Dysphoria Among Transmasculine and Nonbinary Adolescents and Young Adults. *JAMA Pediatr* 2022; **176**(11): 1115-22.
22. Chen D, Berona J, Chan YM, et al. Psychosocial Functioning in Transgender Youth after 2 Years of Hormones. *N Engl J Med* 2023; **388**(3): 240-50.
23. Foster Skewis L, Bretherton I, Leemaqz SY, Zajac JD, Cheung AS. Short-Term Effects of Gender-Affirming Hormone Therapy on Dysphoria and Quality of Life in Transgender Individuals: A Prospective Controlled Study. *Front Endocrinol (Lausanne)* 2021; **12**: 717766.
24. Haraldsen IR, Haug E, Falch J, Egeland T, Opjordsmoen S. Cross-sex pattern of bone mineral density in early onset gender identity disorder. *Horm Behav* 2007; **52**(3): 334-43.
25. Kaltiala R, Heino E, Työläjärvi M, Suomalainen L. Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria. *Nord J Psychiatry* 2020; **74**(3): 213-9.
26. 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.
27. Turan Ş, Aksoy Poyraz C, Usta Sağlam NG, et al. Alterations in Body Uneasiness, Eating Attitudes, and Psychopathology Before and After Cross-Sex Hormonal Treatment in Patients with Female-to-Male Gender Dysphoria. *Arch Sex Behav* 2018; **47**(8): 2349-61.
28. van Dijk D, Dekker M, Conemans EB, et al. Explorative Prospective Evaluation of Short-Term Subjective Effects of Hormonal Treatment in Trans People-Results from the European Network for the Investigation of Gender Incongruence. *J Sex Med* 2019; **16**(8): 1297-309.
29. van Kesteren P, Lips P, Deville W, et al. The effect of one-year cross-sex hormonal treatment on bone metabolism and serum insulin-like growth factor-1 in transsexuals. *J Clin Endocrinol Metab* 1996; **81**(6): 2227-32.
30. Vlot MC, Wiepjes CM, de Jongh RT, T'Sjoen G, Heijboer AC, den Heijer M. Gender-Affirming Hormone Treatment Decreases Bone Turnover in Transwomen and Older Transmen. *J Bone Miner Res* 2019; **34**(10): 1862-72.
31. Wiepjes CM, de Jongh RT, de Blok CJ, et al. Bone Safety During the First Ten Years of Gender-Affirming Hormonal Treatment in Transwomen and Transmen. *J Bone Miner Res* 2019; **34**(3): 447-54.

32. Wiepjes CM, Vlot MC, Klaver M, et al. Bone Mineral Density Increases in Trans Persons After 1 Year of Hormonal Treatment: A Multicenter Prospective Observational Study. *Journal of Bone and Mineral Research* 2017; **32**(6): 1252-60.
33. Nota NM, Wiepjes CM, de Blok CJM, Gooren LJG, Kreukels BPC, den Heijer M. Occurrence of Acute Cardiovascular Events in Transgender Individuals Receiving Hormone Therapy. *Circulation* 2019; **139**(11): 1461-2.
34. Pyra M, Casimiro I, Rusie L, et al. An Observational Study of Hypertension and Thromboembolism Among Transgender Patients Using Gender-Affirming Hormone Therapy. *Transgend Health* 2020; **5**(1): 1-9.
35. Becker-Hebly I, Fahrenkrug S, Campion F, Richter-Appelt H, Schulte-Markwort M, Barkmann C. Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: a descriptive study from the Hamburg Gender Identity Service. *Eur Child Adolesc Psychiatry* 2021; **30**(11): 1755-67.
36. Filipov H, Kavla Y, Şahin S, Gökler ME, Turan Ş. The Effects of Gender-Affirming Hormone Therapy on Body Satisfaction, Self-Esteem, Quality of Life, and Psychopathology in People with Female-to-Male Gender Dysphoria. *Transgend Health* 2023; **8**(2): 168-74.
37. Wells GAW, G.; Shea, B.; Shea, B.; O'Connell, D.; Peterson, J.; Welch, Losos, M.; Tugwell, P.; Ga, S.W.; Zello, G.A.; & Petersen, J.A.;. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. Ottawa Hospital Research Institute. 2014.
38. Brignardello-Petersen R, Guyatt GH. Assessing the Certainty of the Evidence in Systematic Reviews: Importance, Process, and Use. *Am J Epidemiol* 2024.
39. Taylor J, Mitchell A, Hall R, Langton T, Fraser L, Hewitt CE. Masculinising and feminising hormone interventions for adolescents experiencing gender dysphoria or incongruence: a systematic review. *Arch Dis Child* 2024.
40. Doyle DM, Lewis TOG, Barreto M. A systematic review of psychosocial functioning changes after gender-affirming hormone therapy among transgender people. *Nature Human Behaviour* 2023; **7**(8): 1320-31.
41. van Leerdam TR, Zajac JD, Cheung AS. The Effect of Gender-Affirming Hormones on Gender Dysphoria, Quality of Life, and Psychological Functioning in Transgender Individuals: A Systematic Review. *Transgend Health* 2023; **8**(1): 6-21.
42. Chew D, Anderson J, Williams KJ, May T, Pang KC. Hormonal Treatment in Young People With Gender Dysphoria: A Systematic Review. *Pediatrics* 2018; **141**.
43. Connelly PJ, Clark A, Touyz RM, Delles C. Transgender adults, gender-affirming hormone therapy and blood pressure: a systematic review. *Journal of Hypertension* 2021; **39**(2).
44. Schünemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations: The GRADE Working Group; 2013.

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



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

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

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