Mental Wellness

Mental Wellness 2026; 4:42
doi: 10.4081/mw.2026.42

Gender-affirming hormone therapy and suicide-related outcomes in transgender and gender-diverse populations: a systematic review and meta-analysis

Mohsen Khosravi,[1-3] Massimo Tusconi,[4] Isa Multazam Noor,[5] Chou-Yi Hsu,[6] Samir Sahoo,[7] Zafar Aminov,[8] Swati Mishra[9]

[1]Department of Psychiatry, School of Medicine, Zahedan University of Medical Sciences, Iran; [2]Health Promotion Research Center, Zahedan University of Medical Sciences, Iran; [3]Community Nursing Research Center, Zahedan University of Medical Sciences, Iran; [4]Department of Medical Sciences and Public Health, University of Cagliari, Italy; [5]Dr Soeharto Heerdjan Neuropsychiatric Hospital, Jakarta, Indonesia; [6]Thunderbird School of Global Management, Arizona State University, Tempe Campus, Phoenix, Arizona, USA; [7]Department of General Medicine, IMS and SUM Hospital, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India; [8]Department of Public Health and Healthcare Management, Samarkand State Medical University, Samarkand, Uzbekistan; [9]Department of Pharmacology, IMS and SUM Hospital, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India

Abstract

This systematic review and meta-analysis examined the association between gender-affirming hormone therapy (GAHT) and suicide-related outcomes (including suicidal ideation, suicide attempt, and suicide death) among transgender and gender-diverse individuals. It included 8 independent studies with 11 effect sizes. The pooled multilevel estimate was not statistically significant but pointed in a negative direction (log odds ratio = -0.3576; 95% confidence interval: -1.2768 to 0.5617; p=0.4458), suggesting a possible but not definitive protective effect of GAHT against suicidality. There was substantial heterogeneity among studies [Q (10)=170.5309, p<0.0001; σ^2=1.7820; I^2≈94.1%]. Moderator analyses showed that age group was significant [QM (2)=6.9350, p=0.0312], with mixed-age samples differing from adolescent samples, but adult samples did not significantly differ from adolescent samples. Comparator type, study design, and outcome domain were not significant moderators. Sensitivity analyses pointed to influential studies and evidence of funnel asymmetry and small-study effects. The conclusions highlight that while there is a non-significant trend toward a protective effect of GAHT, the evidence is limited by high heterogeneity, influential studies, small-study effects, and the predominantly observational nature of the data. Larger and better-characterized longitudinal studies are needed for clearer causal inference.

Key words: gender-affirming hormone therapy, transgender, suicidality, meta-analysis.

Correspondence to: Mohsen Khosravi, Department of Psychiatry, School of Medicine, Zahedan University of Medical Sciences, Zahedan, Iran.
E-mail: dr_khosravi2016@yahoo.com

Introduction

Transgender and gender-diverse (TGD) individuals (including binary and non-binary identities) experience pronounced health inequities, with disparities in mental health among the most consistent and severe findings in the literature.[1-3] Epidemiological estimates suggest that approximately 0.6% to 1.1% of the general population identify as TGD, although prevalence varies by geographic region, measurement strategy, and definitional criteria.[4,5] Among adolescents, reported prevalence is often higher, with national surveys indicating TGD identification in 2.3% of Australian and 1.2% of New Zealand youth,[6,7] and U.S.-based school surveillance studies reporting comparable rates (1.3-1.8%) alongside additional subgroups expressing gender uncertainty.[8] In parallel, referrals to pediatric gender services have increased substantially over the past decade, likely reflecting greater visibility, evolving social recognition, and expanded access to gender-affirming care pathways.[9-11] The burden of suicidality in TGD populations represents one of the most concerning mental health disparities documented to date. Lifetime suicide attempt prevalence has been estimated to approach 30%, with higher rates reported in some adolescent samples.[12] These elevated risks are increasingly understood not as isolated consequences of medical treatment exposure, but as arising within an interactional framework that includes gender dysphoria, minority stress, structural stigma, victimization, and barriers to affirming care.[13-16] Recent evidence strengthens this interpretation. A systematic review of self-harm thoughts and behaviors in TGD populations found that discrimination, victimization, and proximal minority stressors were consistently associated with suicidal ideation and attempts, whereas affirming social environments functioned as protective factors, while also documenting substantial heterogeneity in measures and study designs.[17] Similarly, a recent meta-analysis of structural, interpersonal, and individual stigma in transgender and nonbinary populations reported robust associations between stigma exposures and adverse health outcomes, while emphasizing considerable between-study heterogeneity and the predominance of cross-sectional evidence.[18] Longitudinal work has further shown that distal and proximal gender minority stressors may differentially relate to internalizing symptoms and suicidality, with hopelessness partially mediating these associations, although causal interpretation remains constrained in baseline-focused analyses.[19] Together, these findings underscore that suicidality in TGD populations is embedded in broader psychosocial and structural determinants. Gender-affirming medical interventions, including puberty suppression with gonadotropin-releasing hormone analogues (GnRHa), gender-affirming hormone therapy (GAHT), and gender-affirming surgeries, are increasingly recognized as medically necessary for many TGD individuals.[20,21] These interventions are intended to reduce distress associated with incongruence between experienced gender and assigned sex at birth and, in appropriate clinical contexts, may contribute to improved psychosocial functioning and mental health.[13,22] International standards emphasize multidisciplinary care integrating medical, mental health, and psychosocial support to optimize outcomes.[20,23] Evidence suggests that the timely initiation of GnRHa may be associated with improved functioning and lower lifetime suicidal ideation in some populations,[24,25] although potential adverse effects (*e.g.*, bone mineral density impacts with prolonged use in some youth populations) must also be considered.[26-28] GAHT, typically involving estradiol with antiandrogens for transfeminine individuals and testosterone for transmasculine individuals, induces substantial somatic changes and has been associated with improvements in depression, anxiety, and general distress in many observational studies.[29-33] Emerging randomized evidence also suggests that earlier GAHT initiation may be associated with rapid reductions in psychological burden, including suicidal ideation, in selected clinical contexts.[34] Gender-affirming surgeries have similarly been associated with improvements in body congruence, quality of life, and sexual well-being, although direct evidence regarding suicide-related mortality remains limited and methodologically complex.[35-39] At the same time, the literature relevant to GAHT and suicide-related outcomes is neither uniform nor methodologically straightforward. Some studies report beneficial associations under specific conditions, whereas others yield null, mixed, or design-sensitive findings, particularly when exposure definitions, comparators, and outcome measures differ. In a prospective cohort of transgender and nonbinary adolescents and young adults attending a multidisciplinary U.S. gender clinic, initiation of puberty blockers or GAHT (*vs.* no initiation) was associated with lower odds of depression and suicidality over 12 months, but the authors explicitly noted the observational design, clinic-based sampling, and potential for residual confounding.[40] In a large U.S. cross-sectional adult survey, access to pubertal suppression among those who desired it was associated with lower lifetime suicidal ideation, but the authors cautioned that recall bias, cohort effects, and the inability to isolate puberty suppression from later interventions limited causal inference.[25] Prospective adult evidence also suggests improvements in psychological functioning following gender-affirming treatment, including hormones, but with heterogeneity across outcome domains, follow-up periods, and baseline symptom severity.[31] Consistent with this, a recent systematic review of prospective adult mental health outcomes following affirmative interventions reported overall improvements in depression and anxiety after hormonal and surgical treatments, while highlighting small samples, lack of control groups, substantial methodological heterogeneity, and limited suicidality-specific evidence.[41] Large-scale observational work further illustrates that inferences about post-intervention mental health outcomes can be sensitive to analytic specification and comparator definition; for example, population-level analyses involving gender-affirming surgical care and mental health treatment utilization have been the subject of methodological scrutiny and interpretation shifts depending on modeling decisions.[42] In youth populations, broader work on gender identity milestones and family support also indicates that affirmation-related mental health outcomes vary substantially by social context, particularly family support, rather than following a uniform trajectory across all adolescents.[43] Collectively, this literature supports the need to examine not only whether associations are observed, but also the conditions under which they strengthen, attenuate, or become difficult to interpret. These inconsistencies are not merely noise; they point to specific methodological and clinical features that may moderate observed associations between GAHT and suicide-related outcomes. Outcome domain is one such source of variation. Risk and protective factors for suicidal ideation, suicide attempts, and related self-harm behaviors do not necessarily operate identically, and many studies in this literature prioritize depressive symptoms or suicidal ideation while attempts and mortality remain comparatively under-examined.[17,41] The developmental stage is another plausible moderator. Adolescence is a sensitive period for identity consolidation, peer stress, and family dynamics, and earlier access to affirming interventions (including puberty suppression and/or hormones) may alter cumulative exposure to minority stress, whereas adult initiation often occurs after prolonged exposure to stigma and distress.[25,41,43] Comparator selection may also materially influence estimates: pre-post clinic-based comparisons may capture symptom change over time but are more vulnerable to expectancy effects, regression to the mean, and concurrent psychosocial support than studies using concurrent non-initiation comparators.[41,42] Study design may likewise influence the magnitude and direction of observed associations through differences in temporal ordering, confounding control, and outcome ascertainment.[17,41] In addition to these primary moderators, sample composition may contribute to heterogeneity because transmasculine, transfeminine, and nonbinary participants are often combined despite potentially different social exposures, baseline risk profiles, and treatment pathways, while stratified analyses remain uncommon.[17-19] Reviews have likewise called for longer follow-up and explicit evaluation of treatment duration to determine whether the mental health correlates of hormone exposure vary over time.[41] Finally, psychosocial moderators such as discrimination, coping style, self-compassion, and community support may buffer or amplify suicide risk independent of, or in interaction with, access to medical care, reinforcing the importance of an integrative analytic framework rather than a treatment-only model.[17,18,44] Despite encouraging findings from some clinical and observational studies, the empirical literature on suicide-related outcomes in relation to GAHT remains fragmented by small samples, variable follow-up periods, inconsistent exposure definitions (*e.g.*, current use, past use, initiation status), heterogeneous comparators (*e.g.*, pre-GAHT baseline *vs.* no-GAHT), and variable operationalization of suicidality (ideation, attempts, and death).[10,41,45,46] Existing reviews often aggregate diverse mental health endpoints or combine distinct gender-affirming interventions, which can obscure intervention-specific patterns for suicide-related outcomes. A rigorous quantitative synthesis focused specifically on GAHT and suicide-related outcomes is therefore warranted, particularly one that evaluates whether associations differ systematically by study design, comparator definition, developmental stage, and outcome domain, while situating those findings within the broader psychosocial context in which suicidality occurs. Accordingly, the present systematic review and meta-analysis aimed to estimate the association between GAHT and suicide-related outcomes (including ideation, attempts, and death) among TGD populations and to examine whether observed effect sizes varied by study design, comparator group, age group, and suicide outcome domain.

Methods

Study design and reporting framework

This systematic review and meta-analysis was conducted to quantitatively synthesize evidence on the association between GAHT and suicide-related outcomes in TGD populations. The review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement,[47] and the literature search was documented in line with the PRISMA-S extension for search reporting.[48]

Protocol and registration

A review protocol was developed prior to study selection and data extraction. The review was not prospectively registered in PROSPERO or another registry. This should be considered a limitation affecting transparency and reproducibility, although the protocol-defined eligibility criteria, extraction fields, and synthesis plan were established before completion of full-text review and meta-analytic modeling.[47]

Information sources and search strategy

A systematic search was conducted in PubMed, PsycINFO, Scopus, and Web of Science from database inception to February 2025. Search strategies combined controlled vocabulary (where available) and free-text terms related to TGD populations, GAHT, and suicide-related outcomes (including suicidal ideation, suicide attempt, and suicide mortality). Reference lists of relevant reviews and eligible empirical studies were hand-searched to identify additional records and recently published studies. The complete database-specific search syntaxes, search dates, database platforms/interfaces, and any applied limits (*e.g.*, language or date) are provided in *Supplementary Material 1* in accordance with PRISMA-S.[48] Grey-literature eligibility (including treatment of preprints, dissertations, conference abstracts, and reports) is also specified in *Supplementary Material 1*.

Record management and duplicate removal

Search results were exported to a reference-management and screening workflow, and duplicates were removed using automated matching (*e.g.*, DOI/title/author matching) followed by manual verification. The record-management and duplicate-removal process is described in *Supplementary Material 1*.[48]

Eligibility criteria

Eligibility criteria were defined *a priori* using a PICOS-informed framework.[47] Eligible studies included TGD participants (including transmasculine, transfeminine, and/or nonbinary participants, as defined by study authors) and examined GAHT as an exposure (current use, initiation, or exposure history). Studies were required to include a comparator condition (*e.g.*, no GAHT, treatment-naïve group, delayed/non-initiation group, or pre-GAHT baseline comparison) and report at least one suicide-related outcome, operationalized as suicidal ideation, suicide attempt, or suicide death/mortality. Eligible study designs were quantitative observational designs (including cross-sectional, prospective cohort, and retrospective cohort studies). Studies were required to provide sufficient statistical information to extract or compute an effect size [log odds ratio (logOR)] and its standard error (SE) [*e.g.*, OR with confidence intervals (CIs) or raw event-count data]. Editorials, commentaries, qualitative studies, case reports, and studies without disaggregated suicide-related outcomes or without sufficient data for effect-size computation were excluded. Language restrictions, publication-type restrictions, and any minimum follow-up or minimum sample-size requirements applied in the review are specified in *Supplementary Material 1*.

Screening and study selection

The study selection process is summarized in the PRISMA flow diagram (Figure 1). The initial search yielded 129 unique records. After title/abstract screening, 41 articles underwent full-text review. Following a full-text eligibility assessment, 8 independent studies contributing 11 effect sizes were retained for meta-analytic synthesis. Title/abstract screening, full-text review, data extraction, and coding were conducted by two reviewers, with disagreements resolved by consensus. Where needed, discrepancies were resolved through re-review of the original report and discussion of coding rules. Reviewer roles at each stage are detailed in *Supplementary Material 2*. Inter-rater agreement was not formally quantified.

Data extraction and coding

A standardized extraction form was used to collect study-level and effect-size-level data. Extracted variables included study identifier, full citation, country, study design, sample size in exposed and comparator/baseline conditions, age group, mean age (when reported), sample composition [including assigned female at birth (AFAB)/assigned male at birth (AMAB) proportions when reported], exposure type, treatment duration (months, when available), comparator type, suicide outcome domain (ideation, attempt, death), outcome timeframe, measurement instrument/ascertainment method, reported effect estimate inputs (*e.g.*, ORs and CIs, or raw counts), computed effect size (logOR), SE, and risk-of-bias classification. Study-level and effect-size-level characteristics used in the synthesis and moderator analyses are presented in Table 1. Where multiple eligible suicide-related outcomes were reported within the same study (*e.g.*, ideation and attempt), each outcome was retained as a separate effect size. Moderator variables were coded using predefined categories: study design (cross-sectional, prospective cohort, retrospective cohort), age group (adolescent, adult, mixed), comparator type (*e.g.*, no GAHT *vs.* pre-GAHT baseline), and suicide outcome domain (ideation, attempt, death). Additional variables, including sample composition and treatment duration, were extracted where reported to assess feasibility for future moderator analyses.

Risk-of-bias assessment

Risk of bias in included non-randomized studies was assessed using the Risk of Bias in Non-randomized Studies of Interventions

(ROBINS-I) framework.[49] ROBINS-I evaluates bias across domains, including confounding, participant selection, classification of interventions, deviations from intended interventions, missing data, outcome measurement, and selection of reported results. Domain-level judgments and overall judgments were recorded for each study and are reported in *Supplementary Material 3*. For consistency of presentation in the main text and Tables 1 and 2, overall judgments were grouped into summary categories used in sensitivity analyses.

Effect-size computation

The primary effect size metric was logOR. When studies reported ORs, effect sizes were transformed using the natural logarithm. SEs were derived from reported CIs using standard meta-analytic conversion procedures (*e.g.*, SE=[ln (upper CI)–ln (upper CI)]/2×1.96)) and/or computed from raw count data when available.[50] Effect directions were harmonized such that negative logOR values indicated lower odds of the suicide-related outcome in the GAHT-exposed group relative to the comparator.

Statistical synthesis

Meta-analytic models were fitted in R (version 4.x) using the metafor package.[51] Because multiple effect sizes were available from some studies, a multilevel random-effects meta-analysis was conducted using rma.mv, consistent with multilevel meta-analytic methods for dependent effect sizes.[52] Random effects were specified at the study level to model dependence among effect sizes from the same study and to estimate between-study heterogeneity. Effect sizes were weighted by inverse variance to derive pooled estimates.

Heterogeneity assessment

Heterogeneity was evaluated using τ^2 (between-study variance), the Q statistic (test of residual heterogeneity), and I^2 (the proportion of total variability attributable to heterogeneity rather than sampling error).[50] Because I^2 estimation in multilevel models can depend on the variance structure, the method used to derive I^2 is reported alongside the results, and I^2 values were interpreted in

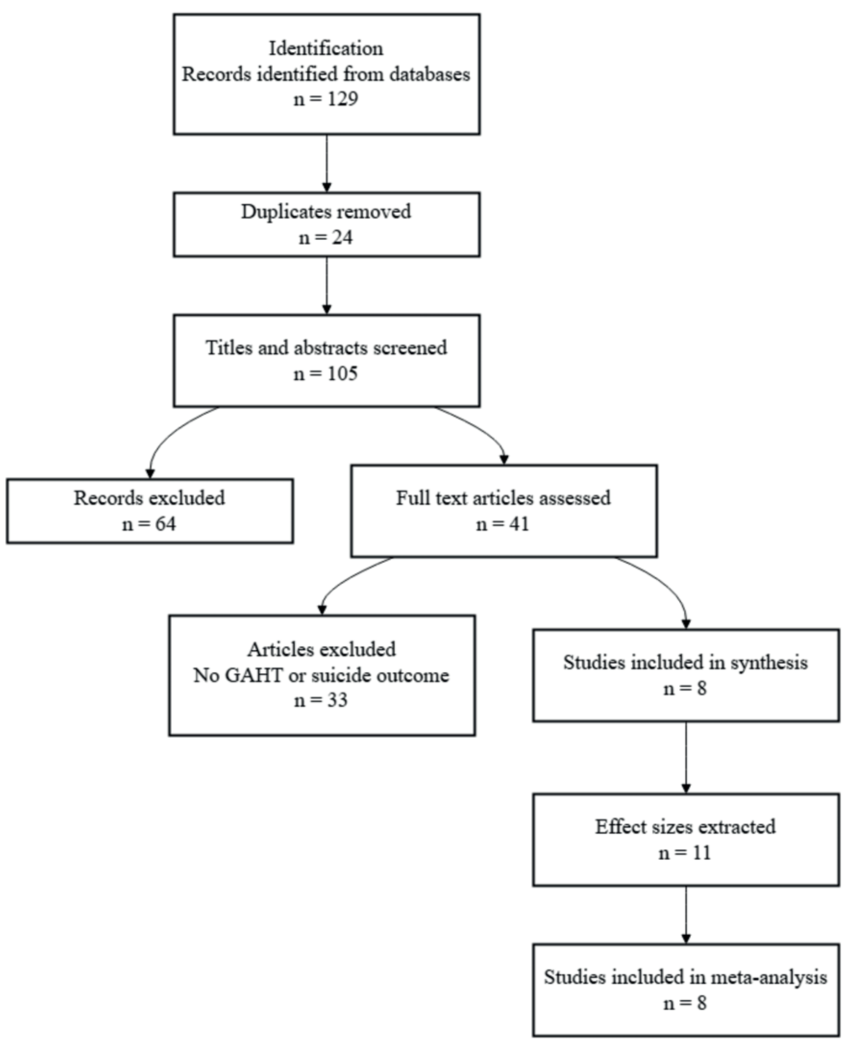

Figure 1. PRISMA flow diagram of study selection. GAHT, gender-affirming hormone therapy.

conjunction with τ^2, Q statistics, and visual dispersion of study estimates.

Moderator analyses

Moderator analyses were conducted using separate multilevel meta-regression models, entering one categorical moderator at a time while retaining the multilevel random-effects structure. Moderators included study design, age group, comparator type, and suicide outcome domain. These variables were selected to evaluate whether effect estimates varied systematically across key methodological and clinical features identified *a priori* in the review rationale.

Bias diagnostics and sensitivity analyses

Several diagnostics were used to examine the robustness of the pooled estimate. Influence diagnostics included leave-one-out analyses and Baujat plots to identify studies contributing disproportionately to the pooled effect size and/or heterogeneity.[53]

Table 1. Characteristics of studies included in the meta-analysis.

Study ID	Year	Country	Study design	Age group	Mean age	N exposed	N comparator	Exposure type	Comparator	Overall risk of bias
Green2022Trevor,[55]	2022	United States	Cross-sectional	Mixed	17.62	1216	4537	GAHT	No GAHT	Some concerns
Green2022Trevor,[55]	2022	United States	Cross-sectional	Mixed	17.62	1216	4537	GAHT	No GAHT	Some concerns
Hughto2020_TSHS,[56]	2020	United States	Cross-sectional	Mixed	32.80	288	288	GAHT	Pre-GAHT baseline	Some concerns
Hughto2020_TSHS_attempt,[56]	2020	United States	Cross-sectional	Mixed	32.80	288	288	GAHT	Pre-GAHT baseline	Some concerns
Turban2020StanfordPeds,[25]	2020	United States	Prospective cohort	Adolescent	15.80	240	112	GAHT	No GAHT	Low
Tordoff2022Seattle,[40]	2022	United States	Prospective cohort	Adolescent	15.80	63	42	GAHT	No GAHT	Some concerns
Summers2021Memphis,[57]	2021	United States	Retrospective cohort	Adult	NR	31	28	GAHT	No GAHT	Some concerns
Colizzi2014-GermanCohort,[58]	2014	Germany	Prospective cohort	Adult	36.20	47	0	GAHT	Pre-GAHT baseline	Some concerns
Bränström2020,[42]	2023	United States	Cross-sectional	Adult	30.20	2454	301	GAHT	No GAHT	Some concerns
Bränström2020,[42]	2023	United States	Cross-sectional	Adult	30.20	2454	301	GAHT	No GAHT	Some concerns
Tucker2018_TransgenderVeterans,[59]	2018	United States	Cross-sectional	Adult	56.50	28	36	GAHT + other intervention	No GAHT	Some concerns

GAHT, gender-affirming hormone therapy.

Table 2. Characteristics of studies included in the meta-analysis.

Study ID	Outcome domain	Outcome timeframe/follow-up	Duration (months)	Outcome measure/ascertainment	logOR	SE
Green2022Trevor,[55]	Ideation	Past year	NR	YRBS	-0.30	0.09
Green2022Trevor,[55]	Attempt	Past year	NR	YRBS	-0.17	0.12
Hughto2020_TSHS,[56]	Ideation	Lifetime	NR	Self-report	1.35	0.19
Hughto2020_TSHS_attempt,[56]	Attempt	Lifetime	NR	Self-report	1.71	0.24
Turban2020StanfordPeds,[25]	Ideation	Past year	12	PHQ-9 item 9	-1.11	0.28
Tordoff2022Seattle,[40]	Attempt	Past year	12	Self-report	-1.20	0.52
Summers2021Memphis,[57]	Death	During follow-up	36	Viral load suppression definition	0.67	0.52
Colizzi2015-GermanCohort,[58]	Ideation	During follow-up	12	Structured self-report interview	-1.83	0.50
Bränström2020,[42]	Ideation	Past year	NR	NR	-0.69	0.28
Bränström2020,[42]	Attempt	Past year	NR	NR	-0.92	0.35
Tucker2018_TransgenderVeterans,[59]	Ideation	Past 2 weeks	NR	PHQ-9 item 9	-2.84	1.07

YRBS, Youth Risk Behavior Survey; PHQ-9, Patient Health Questionnaire-9; NR, not reported.

Publication bias/small-study effects were assessed through visual inspection of funnel plot asymmetry and formally tested using Egger's regression test.[54] Because the primary synthesis used a multilevel model and formal asymmetry testing was implemented in a single-level framework, Egger's test was conducted as a random-effects sensitivity analysis using *rma*. Funnel-plot asymmetry and Egger's test were interpreted cautiously, given the small number of included effect sizes and substantial heterogeneity.[48] Additional sensitivity analyses included subgroup analyses by risk-of-bias classification and outlier-influence checks comparing pooled estimates with and without influential studies.

Data presentation and visualization

Forest plots, funnel plots, and influence diagnostics (including Baujat plots) were generated in R, including ggplot2-based visualizations. Figures were exported in high-resolution formats suitable for publication. Table 1 and Table 2 present the expanded study and effect-size characteristics (including sample size, country, age group, comparator type, outcome domain, outcome timeframe, extracted duration where available, and effect-size inputs/outputs) used in the primary and moderator analyses.[25,40,42,55-59]

Results

This meta-analysis synthesized 11 effect sizes drawn from 8 independent studies examining the association between GAHT and suicide-related outcomes among TGD populations. All effect sizes were expressed as logORs, with negative values indicating lower odds of the suicide-related outcome (suicidal ideation, suicide attempt, or suicide death) in the GAHT-exposed group relative to the comparator. In the multilevel random-effects meta-analysis, the pooled estimate was logOR=-0.3576 (SE=0.4690; 95% CI: -1.2768 to 0.5617; p=0.4458), indicating no statistically significant overall association in the pooled model. Heterogeneity was substantial [Q (10)=170.5309, p<0.0001], and the estimated between-study variance component in the multilevel model was $\sigma^2=1.7820$ (τ^2-equivalent variance component). The corresponding overall heterogeneity was high ($I^2\approx94.1\%$, based on the Q-statistic approximation; see Methods for reporting approach in the multilevel setting). The study-level effect estimates varied in both direction and magnitude (Figure 2). Observed logORs ranged from -2.84 to 1.71. Statistically significant negative estimates (95% CI excluding 0) were observed for several effect sizes, including Turban2020StanfordPeds (ideation), Tordoff2022Seattle (attempt), Colizzi2015-GermanCohort (ideation), Bränström 2020 (ideation and attempt), Green2022Trevor (ideation), and Tucker2018_TransgenderVeterans (ideation). The Green2022Trevor (attempt) estimate was negative but not statistically significant. Positive estimates were observed for 2020_TSHS (ideation) and Hughto2020_TSHS_attempt (attempt), with the attempt estimate excluding 0 and the ideation estimate remaining close to the null boundary. The single death-outcome estimate (Summers2021Memphis) was positive and imprecise, with a CI crossing 0. Figure 2 presents the multilevel forest plot of all included effect sizes and the pooled random-effects estimate.

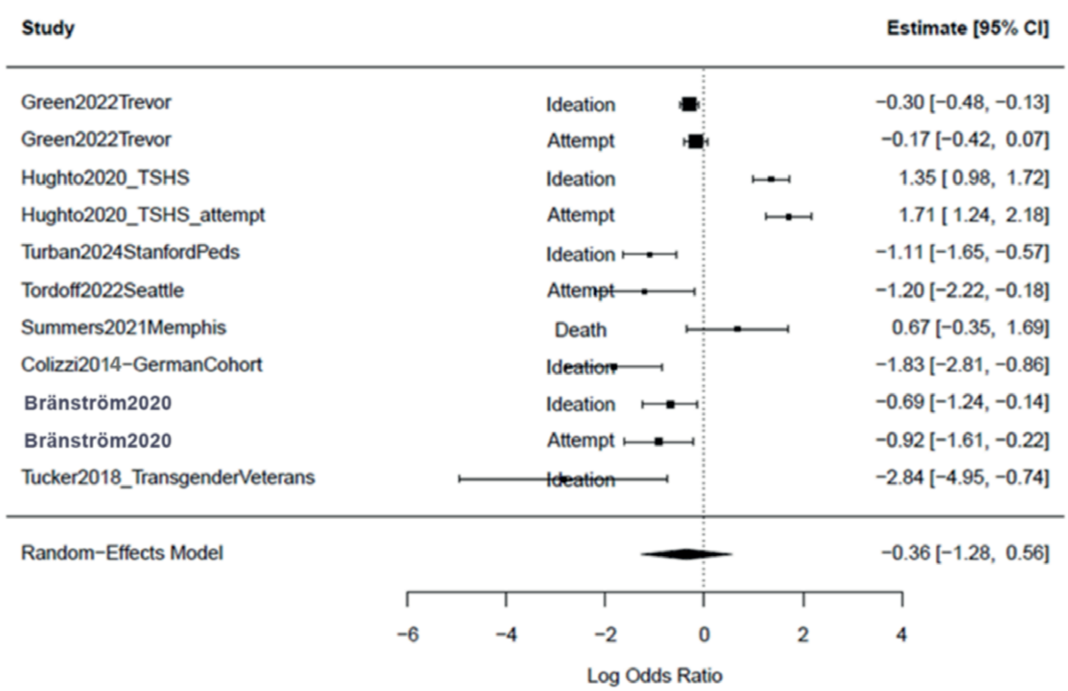

Figure 2. Forest plot of associations between gender-affirming hormone therapy (GAHT) and suicide-Related outcomes. Forest plot displaying study-level log odds ratios and 95% confidence intervals (CIs) for the association between GAHT and suicide-related outcomes. Negative values indicate lower odds of the suicide-related outcome in the GAHT-exposed group relative to the comparator. Squares represent point estimates and horizontal lines represent 95% CIs; square size reflects inverse-variance weighting. The diamond denotes the pooled estimate from the multilevel random-effects model and its 95% CI.

Heterogeneity and moderator analyses

Substantial heterogeneity was observed in the pooled association between GAHT and suicide-related outcomes. In the primary multilevel random-effects model, heterogeneity was statistically significant [Q (10)=170.5309, p<0.0001], with an estimated between-study variance component of σ^2=1.7820 (τ^2-equivalent variance component in the multilevel model). The corresponding overall heterogeneity was high (I^2≈94.1%, based on the Q-statistic approximation; see *Methods* for the reporting approach used in the multilevel setting). Across moderator models, tests of residual heterogeneity remained statistically significant (Q_E range=68.3119 to 164.6506, all p<0.001), indicating that substantial variability in effect sizes persisted after inclusion of individual moderators. To examine potential sources of heterogeneity, 4 prespecified moderators were evaluated using separate multilevel meta-regression models with random intercepts at the study level: age group, comparator type, study design, and suicide outcome domain. Age group was the only moderator with a statistically significant omnibus test [Q_M(2)=6.9350, p=0.0312; Q_E(8)=1 30.5838, p<0.0001; σ^2=1.0144]. In this model, the reference category was adolescent samples. The intercept estimate for the adolescent reference group was logOR=-1.1524 (p=0.1335). The coefficient for adult samples was b=0.1518 (p=0.8747), and the coefficient for mixed-age samples was b=2.0693 (p=0.0327), indicating a statistically significant difference between mixed-age and adolescent samples in the fitted model. Comparator type did not significantly moderate effect sizes [Q_M(1)=1.9721, p=0.1602; Q_E(9)=68.3119, p<0.0001; σ^2=1.5449]. In the fitted model, the reference category was the no-GAHT comparator. The intercept estimate was logOR=-0.7989 (p=0.1430), and the coefficient for pre-GAHT comparator studies was b=1.2935 (p=0.1602). Study design was also not a statistically significant moderator [Q_M(2)=3.0930, p=0.2130; Q_E(8)=1 37.3456, p<0.0001; σ^2=1.4498]. With cross-sectional studies as the reference category, the intercept estimate was logOR=0.0824 (p=0.8851). The coefficient for prospective cohort studies was b=-1.4551 (p=0.1193), and the coefficient for retrospective cohort studies was b=0.5909 (p=0.6796). Suicide outcome domain did not significantly moderate effect sizes [Q_M(2)=1.1001, p=0.5769; Q_E(8)=164.6506, p<0.0001; σ^2=1.8936]. In this model, suicide attempt was the reference category. The intercept estimate was logOR=-0.4188 (p=0.4196). Relative to attempts, the coefficient for suicidal ideation was b=-0.1058 (p=0.4604), and the coefficient for suicide death was b=1.0921 (p=0.4840). The suicide death category was represented by a single effect size in the dataset. Influence diagnostics were examined to evaluate contributions to heterogeneity and model influence. The Baujat plot (Figure 3) identified Hughto *et al*. and Tucker as influential observations with high contributions to heterogeneity and/or influence on the pooled estimate.[56] Results of outlier- and influence-based sensitivity analyses are reported in the sensitivity analyses subsection.

Sensitivity and publication bias analyses

To examine the robustness of the pooled estimate and assess possible funnel plot asymmetry/small-study effects, diagnostic analyses included visual inspection of the forest plot (Figure 4) and funnel plot (Figure 5), together with Egger's regression test. As described in the *Methods*, because the primary synthesis used a multilevel model, Egger's test was conducted as a single-level random-effects sensitivity analysis. Egger's regression test indicated statistically significant funnel plot asymmetry (z=-2.0723, p=0.0382). The corresponding limit estimate (as SE approaches 0) was b=0.6907 (95% CI: -0.4903 to 1.8716).

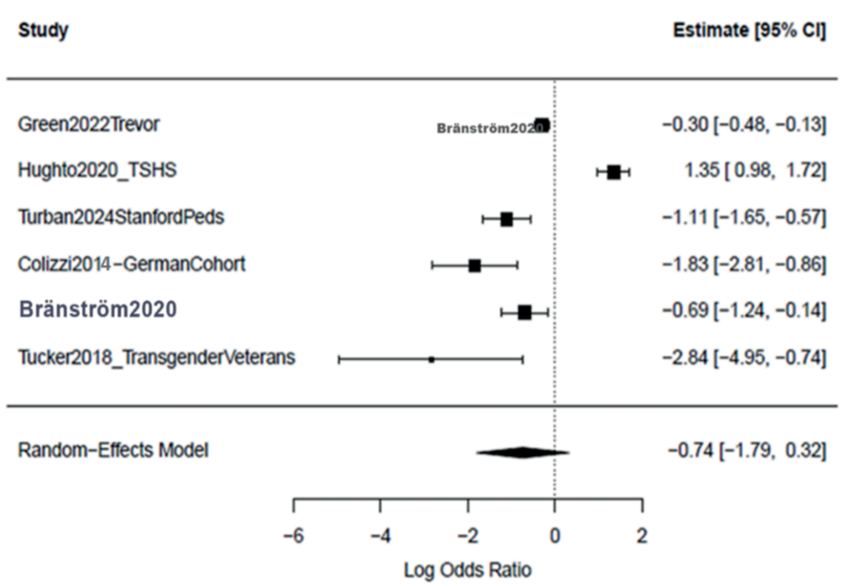

Figure 3. Baujat plot for influence and heterogeneity contribution across included effect sizes-studies. Baujat plot displaying each included effect size-study according to its contribution to total heterogeneity (Q; y-axis) and its influence on the pooled log odds ratio estimate (x-axis). Points located toward the upper-right portion of the plot indicate greater contribution to heterogeneity and/or pooled-effect influence. CI, confidence interval.

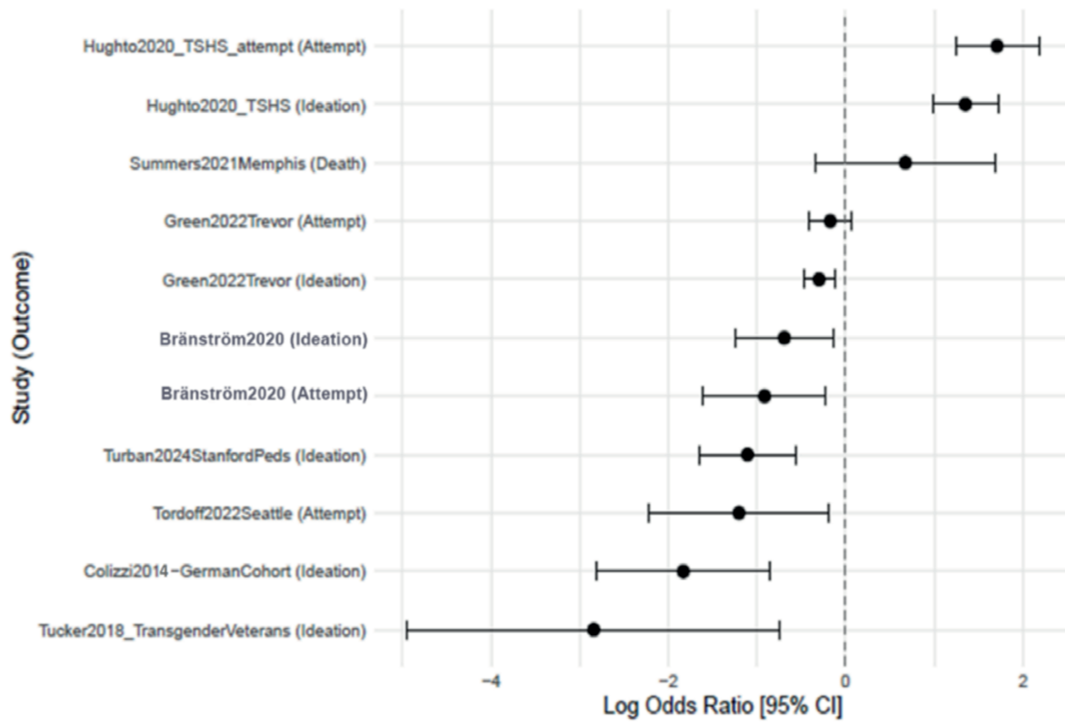

Figure 4. Forest plot of study-level log odds ratios (logORs) for gender-affirming hormone therapy (GAHT) and suicide-related outcomes. Forest plot displaying study-specific logORs and 95% confidence intervals (CIs) for the association between GAHT and suicide-related outcomes. Negative values indicate lower odds of the outcome in the GAHT-exposed group relative to the comparator, and positive values indicate higher odds. The dashed vertical line marks the null value (logOR=0).

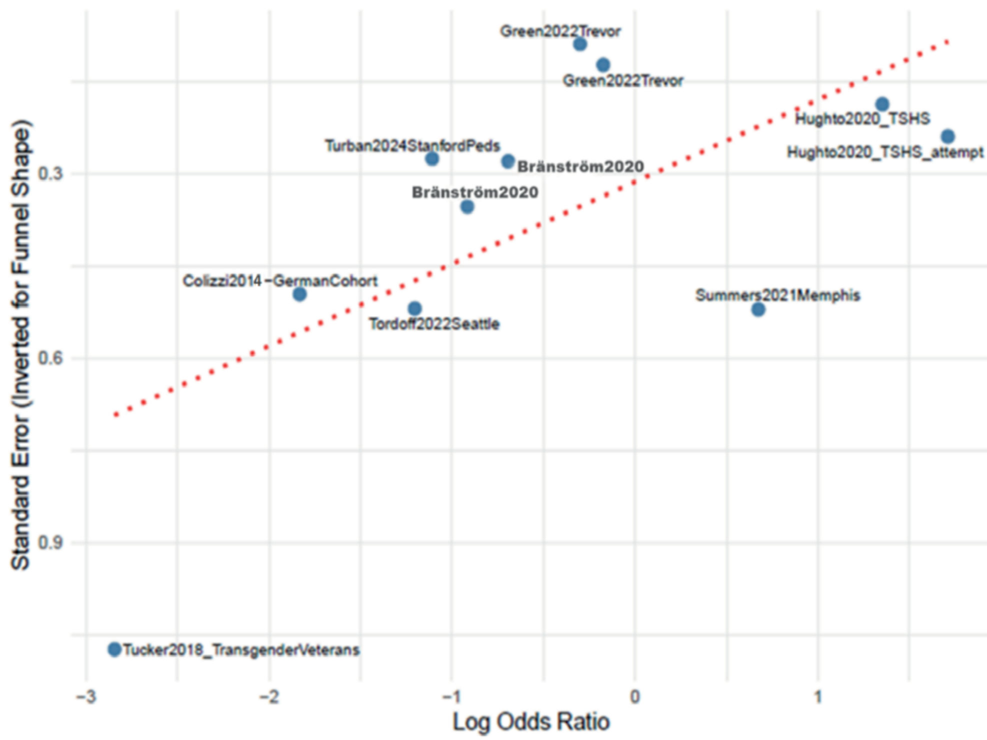

Figure 5. Funnel plot of log odds ratios for assessment of asymmetry/small-study effects. Funnel plot of study-level log odds ratios plotted against standard error (inverted on the y-axis). Labels identify individual study/effect rows. Visual asymmetry was evaluated in conjunction with Egger's regression test, which was conducted as a single-level random-effects sensitivity analysis.

These results were considered evidence of funnel asymmetry/small-study effects in the included effect sizes. Given the small number of effect sizes (k=11), the multilevel dependence structure of the primary dataset, and the substantial heterogeneity observed in the main model, these findings were interpreted cautiously. Visual inspection of the forest plot (Figure 4) showed substantial dispersion in study-level estimates, with effect sizes distributed on both sides of the null and with varying precision. The most negative estimate was observed for Tucker2018_TransgenderVeterans (Ideation) (logOR=-2.84, 95% CI: -4.95 to -0.74), whereas the most positive estimate was observed for Hughto2020_TSHS_attempt (Attempt) (logOR=1.71, 95% CI: 1.24 to 2.18). Additional positive estimates were observed for Hughto2020_TSHS (Ideation) and Summers2021Memphis (Death), with the latter showing a CI crossing 0. This distribution was consistent with the heterogeneity reported in the primary and moderator models. Visual inspection of the funnel plot (Figure 5) showed an uneven distribution of study-level estimates, including low-precision points at more extreme negative values and positive estimates among more precise observations. In particular, Tucker2018_TransgenderVeterans appeared as a low-precision, strongly negative estimate, and the Hughto2020_TSHS and Hughto2020_TSHS_attempt rows appeared on the positive side of the distribution. These visual features were considered alongside the Egger test result as part of the assessment of asymmetry/small-study effects. Formal influence diagnostics (including Baujat plotting; Figure 3) identified the Hughto *et al*. rows and Tucker as influential observations.[56] Exclusion-based sensitivity analyses were therefore conducted by re-estimating the multilevel pooled model after removing influential publications/effect sets. In publication-level sensitivity analyses, exclusion of the two Hughto *et al*.[56] effect sizes yielded a pooled estimate of logOR=-0.8720 (95% CI: -1.5119 to -0.2321) and reduced heterogeneity (σ^2=0.5475; I^2≈76.8%). Exclusion of Tucker alone yielded logOR=-0.4015 (95% CI: -1.2622 to 0.4591), with heterogeneity remaining high (σ^2=1.2249; I^2≈94.5%).[56] Exclusion of both the Hughto *et al*. rows and Tucker yielded logOR=-0.7368 (95% CI: -1.3373 to -0.1364), with reduced heterogeneity (σ^2=0.4321; I^2≈76.1%).[56] These exclusion-based sensitivity analyses showed changes in both the pooled estimate and heterogeneity after removal of influential studies with extreme and divergent effect estimates. Sensitivity analyses stratified by risk-of-bias classification were limited by sparse subgroup counts. Of the 11 effect sizes, 10 were classified as "some concerns" and 1 as "low" risk of bias, precluding a meaningful pooled subgroup comparison for the low-risk category. A restricted sensitivity analysis including only effect sizes rated "some concerns" yielded a pooled estimate of logOR=-0.5335 (95% CI: -1.5500 to 0.4831; p=0.3037), with substantial heterogeneity remaining (σ^2=1.6495; I^2≈94.2%). The single low-risk effect (Turban2020StanfordPeds, ideation) was negative (logOR=-1.1087) and is reported descriptively.

Discussion

With respect to the primary study aim, this systematic review and multilevel meta-analysis did not identify a statistically significant pooled association between GAHT and suicide-related outcomes in TGD populations. The pooled estimate was directionally negative, but the CI was wide and crossed the null, indicating substantial uncertainty in the magnitude and direction of the overall association. This pattern is more appropriately interpreted as a non-significant, directionally protective trend rather than evidence of a definitive protective effect. Importantly, heterogeneity was very high, and study-level estimates varied markedly in both direction and precision, indicating that the included effects are unlikely to reflect a single, stable underlying association across all study contexts. This overall pattern is consistent with the rationale articulated in the *Introduction*: the literature on GAHT and suicide-related outcomes is methodologically heterogeneous and clinically context-dependent, rather than uniformly positive or negative. Prior work has documented substantial heterogeneity across mental health outcomes, study designs, and follow-up intervals in TGD populations.[17,18,41] In addition, suicidality in TGD populations is embedded within broader psychosocial and structural determinants, including minority stress, discrimination, victimization, barriers to care, and social support rather than being reducible to treatment exposure alone.[14-17] The mixed pattern observed here, including both directionally protective and directionally adverse estimates, is therefore compatible with a multifactorial model in which the observed association between GAHT and suicide-related outcomes may vary depending on developmental stage, comparator definition, baseline distress, social context, and co-occurring supports. With respect to the secondary aim (moderation), age group was the only prespecified moderator with a statistically significant omnibus test in the reported models. In the fitted model, mixed-age samples differed significantly from adolescent samples, whereas adult samples did not significantly differ from adolescent samples. At the same time, the adolescent reference estimate itself was not statistically significant, and residual heterogeneity remained substantial after inclusion of age group. Accordingly, the age-group finding should be interpreted as an important but exploratory signal rather than definitive evidence that GAHT has stronger suicide-related effects in adolescents than in adults. A more precise interpretation is that developmental stage may contribute to effect-size variability, and that combining adolescents and adults within the same analytic strata may attenuate or obscure clinically meaningful patterns. This cautious interpretation is nevertheless consistent with prior literature suggesting that developmental timing, family context, and cumulative exposure to minority stress may shape mental health trajectories in TGD youth and young adults.[24,25,34,43] The remaining prespecified moderators (comparator type, study design, and suicide outcome domain) did not significantly explain effect-size variability in the present analyses. However, the non-significant findings should not be interpreted as evidence that these factors are unimportant. Rather, they likely reflect a combination of limited statistical power, sparse subgroup counts, imprecise estimates, and persistent between-study heterogeneity. This is especially relevant for comparator type and study design, which were identified *a priori* as conceptually important sources of variation. As discussed in prior methodological critiques and longitudinal reviews, differences in comparator definitions (*e.g*., pre-GAHT baseline *vs.* no-GAHT comparators) and analytic specification can materially influence apparent treatment associations in gender-affirming care research.[41,42] Similarly, the non-significant outcome-domain moderator should be interpreted cautiously, particularly because the suicide death category was represented by a single effect size and is therefore too sparse to support strong inferences. Influence diagnostics and exclusion-based sensitivity analyses provide further context for interpreting the pooled estimate. The Baujat plot identified the Hughto *et al.* rows and Tucker as influential observations,[56] and re-estimating the pooled model after excluding these influential effect sets changed both the pooled estimate and the heterogeneity magnitude. In particular, exclusion of the Hughto rows and exclusion of both Hughto and Tucker shifted the pooled estimate further in the negative direction and reduced heterogeneity, whereas exclusion of Tucker alone produced a more modest change, and heterogeneity remained high. These results indicate that the pooled association is sensitive to influential studies with extreme and divergent estimates, reinforcing the need to avoid overgeneralizing from the primary pooled model alone. At the same time, these analyses should be viewed as robustness checks rather than replacements for the primary analysis. The asymmetry diagnostics likewise support a cautious interpretation. Visual inspection of the funnel plot suggested an uneven distribution of effects, and Egger's regression indicated statistically significant asymmetry in the single-level sensitivity analysis. However, this should be interpreted as evidence of funnel asymmetry-small-study effects, not as definitive proof of publication bias. In the present dataset, asymmetry may plausibly reflect several overlapping mechanisms, including selective reporting, sparse small-study evidence, heterogeneity-driven asymmetry, possible publication/reporting bias, and dependence among effect sizes. This is particularly important in a small evidence base with widely varying study designs, comparator conditions, and outcome definitions. Accordingly, the asymmetry findings strengthen the case for caution but do not by themselves resolve the direction of the underlying association. Taken together, the present findings are best understood as consistent with a broader, integrative account of suicide risk in TGD populations, in which access to GAHT may be one relevant component of care for some individuals, but not a stand-alone determinant of suicide-related outcomes. This interpretation is congruent with minority stress-informed frameworks and with recent evidence emphasizing the roles of stigma, discrimination, social support, and other psychosocial factors in shaping mental health outcomes.[14,17-19] It is also consistent with evidence that mental health trajectories following gender-affirming interventions vary by social context (*e.g.*, family support), baseline symptom burden, and care environment.[41,43]

From a clinical perspective, these results should be interpreted alongside (not in place of) existing clinical guidelines and individualized care standards.[20,21,23] The present meta-analysis does not establish a uniform causal effect of GAHT on suicide-related outcomes across all TGD populations. However, neither does it support a simplistic conclusion that associations are absent or clinically irrelevant in all settings. Rather, the findings underscore the importance of individualized, multidisciplinary care models that integrate gender-affirming medical treatment with mental health support, family and social interventions, and efforts to reduce structural and interpersonal barriers to care.[13,20] Given the age-related moderation signal observed here, developmental context may be especially important to consider in future clinical and research work, while recognizing that the present age-group findings remain exploratory. Several limitations should be explicitly acknowledged. First, the evidence base was small (8 studies; 11 effect sizes), limiting precision and reducing power for moderator analyses. Second, heterogeneity was substantial in the primary model and remained significant across all moderator models, indicating that important sources of variation were not captured by the prespecified moderators. Third, the included evidence was predominantly observational and non-randomized, with associated risks of residual confounding, confounding by indication, reverse causality (particularly in cross-sectional designs), and comparator non-equivalence. Fourth, outcome ascertainment and exposure definitions varied substantially across studies, including differences in suicidality operationalization (ideation, attempts, death), timeframe, and GAHT exposure definition (*e.g.*, initiation, current use, history), which limits comparability. Fifth, evidence for suicide mortality was extremely sparse (one effect size), precluding meaningful domain-specific conclusions for death outcomes. Sixth, asymmetry/small-study effects were detected but are difficult to attribute definitively in a dataset of this size and structure; possible publication/reporting bias remains a concern but cannot be isolated from other sources of asymmetry. Seventh, risk-of-bias sensitivity analyses were constrained by imbalanced subgroup counts (predominantly "some concerns", with only one low-risk effect), limiting inference from risk-stratified comparisons. Eighth, as a systematic review, the protocol was developed *a priori* but was not prospectively registered, which may reduce transparency and reproducibility relative to prospectively registered evidence syntheses. An additional limitation, also relevant to future research priorities, is that several clinically plausible moderators could not be evaluated robustly in the present meta-regression framework because of incomplete or non-comparable reporting across studies. These include sample composition (*e.g.*, transmasculine, transfeminine, and nonbinary subgroup proportions; AFAB/AMAB composition), treatment duration and time since GAHT initiation, and psychosocial-contextual variables such as discrimination exposure, family support, coping style, self-compassion, and community connectedness.[17,18,41,44] Given the conceptual importance of these variables, future syntheses will depend on more standardized primary-study reporting and more consistent subgroup-specific analyses. Despite these limitations, the study has several strengths. It focuses specifically on GAHT and suicide-related outcomes (rather than aggregating across broader mental health endpoints or multiple intervention types), applies a multilevel meta-analytic framework to account for dependent effect sizes, and evaluates prespecified moderators together with influence and asymmetry diagnostics. This approach provides a more differentiated picture of the current evidence base than a single pooled estimate alone. Overall, the present findings support a cautious and context-sensitive interpretation. The pooled association between GAHT and suicide-related outcomes was non-significant and highly heterogeneous. Age group emerged as the only statistically significant moderator in the reported models, and both asymmetry diagnostics and influence analyses indicated that the summary estimate is sensitive to study-level features. These results reinforce the need for larger, better-characterized prospective studies with clearer comparator definitions, standardized suicidality measures, longer follow-up, and explicit reporting of treatment duration, subgroup composition, and psychosocial context.

Conclusions

This systematic review and multilevel meta-analysis provides a focused quantitative synthesis of the association between GAHT and suicide-related outcomes in TGD populations. The pooled association was non-significant and directionally negative, and therefore is most appropriately interpreted as a potential protective trend rather than definitive evidence of a protective effect. At the same time, the evidence base was highly heterogeneous, and the summary estimate was sensitive to influential studies and asymmetry/small-study effects, warranting caution in drawing generalized conclusions. Among the prespecified moderators, age group was the only statistically significant source of effect-size variation; specifically, mixed-age samples differed from adolescent samples, whereas adult samples did not significantly differ from adolescent samples in the fitted model. This pattern suggests that the developmental stage may contribute to heterogeneity, but it should be interpreted as exploratory given the small number of studies, sparse subgroup structure, and substantial residual heterogeneity. Comparator type, study design, and suicide outcome domain did not significantly moderate effects in the present analyses, although these factors remain clinically and methodologically important and may have been underpowered in the available dataset. The findings are consistent with current clinical guidance that situates GAHT within individualized, multidisciplinary gender-affirming care, while also underscoring that suicide-related outcomes in TGD populations are shaped by broader psychosocial and structural determinants in addition to medical treatment access. The present evidence does not support uniform causal claims across all TGD populations or study contexts. Future research should prioritize larger and better-characterized prospective studies, clearer and more comparable comparator definitions, standardized and explicitly reported suicide-related outcomes (including attempts and mortality), and longer follow-up periods. More consistent reporting is also needed for clinically relevant moderators that could not be robustly tested here, including treatment duration/time since GAHT initiation, sample composition (*e.g.*, transmasculine, transfeminine, and nonbinary subgroup proportions), and psychosocial-contextual factors such as discrimination exposure, family support, and community connectedness. In a field marked by both clinical urgency and sociopolitical salience, methodologically rigorous, transparently reported, and analytically cautious research remains essential for informing care, policy, and suicide prevention efforts among TGD populations.

References

1. Bretherton I, Thrower E, Zwickl S, et al. The health and well-being of transgender Australians: a national community survey. LGBT Health 2021;8:42-9.
2. Clements-Nolle K, Marx R, Katz M. Attempted suicide among transgender persons: The influence of gender-based discrimination and victimization. J Homosex 2006;51:53-69.
3. Maguen S, Shipherd JC. Suicide risk among transgender individuals. Psychology Sexuality 2010;1:34-43.
4. Herman JL, Flores AR, Brown TN, et al. Age of individuals who identify as transgender in the United States. Los Angeles, CA, USA: The Williams Institute; 2017.
5. T'Sjoen G, Arcelus J, Gooren L, et al. Endocrinology of transgender medicine. Endocr Rev 2019;40:97-117.
6. Clark TC, Lucassen MF, Bullen P, et al. The health and well-being of transgender high school students: results from the New Zealand adolescent health survey (Youth'12). J Adolesc Health 2014;55:93-9.
7. Strauss P, Cook A, Winter S, et al. Trans pathways: the mental health experiences and care pathways of trans young people. Summary of results. Perth, Australia: Telethon Kids Institute; 2017
8. Johns MM, Lowry R, Andrzejewski J, et al. Transgender identity and experiences of violence victimization, substance use, suicide risk, and sexual risk behaviors among high school students—19 states and large urban school districts, 2017. MMWR Morb Mortal Wkly Rep 2019;68:67-71.
9. Kaltiala-Heino R, Lindberg N. Gender identities in adolescent population: methodological issues and prevalence across age groups. Eur Psychiatry 2019;55:61-6.
10. White AA, Lin A, Bickendorf X, et al. Potential immunological effects of gender-affirming hormone therapy in transgender people–an unexplored area of research. Ther Adv Endocrinol Metab 2022;13:20420188221139612.
11. Wiepjes CM, Nota NM, de Blok CJ, et al. The Amsterdam cohort of gender dysphoria study (1972–2015): trends in prevalence, treatment, and regrets. J Sex Med 2018;15:582-90.
12. Adams N, Hitomi M, Moody C. Varied reports of adult transgender suicidality: Synthesizing and describing the peer-reviewed and gray literature. Transgend Health 2017;2:60-75.
13. Rosenthal SM. Challenges in the care of transgender and gender-diverse youth: an endocrinologist's view. Nat Rev Endocrinol 2021;17:581-91.
14. Bränström R, Stormbom I, Bergendal M, Pachankis JE. Transgender□based disparities in suicidality: A population□based study of key predictions from four theoretical models. Suicide Life Threat Behav 2022;52:401-12.
15. Testa RJ, Michaels MS, Bliss W, et al. Suicidal ideation in transgender people: Gender minority stress and interpersonal theory factors. J Abnorm Psychol 2017;126:125-36.
16. Romanelli M, Lu W, Lindsey MA. Examining mechanisms and moderators of the relationship between discriminatory health care encounters and attempted suicide among US transgender help-seekers. Adm Policy Ment Health 2018;45: 831-49.
17. Bird K, Arcelus J, Matsagoura L, et al. Risk and protective factors for self-harm thoughts and behaviours in transgender and gender diverse people: a systematic review. Heliyon 2024;10:e26074.
18. Huynh KD, Lefevor GT, Matsuno E, et al. A systematic review and meta-analysis of the associations between structural, interpersonal, and individual stigma and health outcomes for transgender and nonbinary people. 2025. Available from: https://www.researchgate.net/profile/Kiet-Huynh/publication/395379979_A_Systematic_Review_and_Meta-Analysis_of_the_Associations_Between_Structural_Interpersonal_and_Individual_Stigma_and_Health_Outcomes_for_Transgender_and_Nonbinary_People/links/68c479e7a14ac57c39a7b8f7/A-Systematic-Review-and-Meta-Analysis-of-the-Associations-Between-Structural-Interpersonal-and-Individual-Stigma-and-Health-Outcomes-for-Transgender-and-Nonbinary-People.pdf. Accessed on: 28/10/2025.
19. Ünsal BC, Demetrovics Z, Reinhardt M. Gender minority stressors, hopelessness, and their associations with internalizing and externalizing mental health outcomes in a Hungarian trans adult sample. Arch Sex Behav 2025;54:1859-74.
20. Coleman E, Radix AE, Bouman WP, et al. Standards of care for the health of transgender and gender diverse people, version 8. Int J Transgend Health 2022;23:S1-259.
21. Hembree WC, Cohen-Kettenis PT, Gooren L, et al. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2017;102:3869-903.
22. Budge SL, Adelson JL, Howard KA. Anxiety and depression in transgender individuals: the roles of transition status, loss,

social support, and coping. J Consult Clin Psychol 2013;81:545-57.
23. World Professional Association for Transgender Health. Standards of care for the health of transsexual, transgender, and gender nonconforming people. 7th ed. 2011. Available from: https://www.cpath.ca/wp-content/uploads/2009/12/Standards-of-Care-V7-2011-WPATH.pdf
24. de Vries AL, Steensma TD, Doreleijers TA, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. J Sex Med 2011;8:2276-83.
25. Turban JL, King D, Carswell JM, Keuroghlian AS. Pubertal Suppression for Transgender Youth and Risk of Suicidal Ideation. Pediatrics 2020;145:e20191725.
26. Ciancia S, Dubois V, Cools M. Impact of gender-affirming treatment on bone health in transgender and gender diverse youth. Endocr Connect 2022;11:e220280.
27. Klink D, Caris M, Heijboer A, et al. Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. J Clin Endocrinol Metab 2015;100:E270-5.
28. Schagen SEE, Wouters FM, Cohen-Kettenis PT, et al. Bone development in transgender adolescents treated with GnRH analogues and subsequent gender-affirming hormones. J Clin Endocrinol Metab 2020;105:e4252-63.
29. Irwig MS. Testosterone therapy for transgender men. Lancet Diabetes Endocrinol 2017;5:301-11.
30. Tangpricha V, den Heijer M. Oestrogen and anti-androgen therapy for transgender women. Lancet Diabetes Endocrinol 2017;5:291-300.
31. Baker KE, Wilson LM, Sharma R, et al. Hormone therapy, mental health, and quality of life among transgender people: a systematic review. J Endocr Soc 2021;5:bvab011.
32. Costa R, Colizzi M. The effect of cross-sex hormonal treatment on gender dysphoria individuals' mental health: a systematic review. Neuropsychiatr Dis Treat 2016;12:1953-66.
33. White Hughto JM, Reisner SL. A systematic review of the effects of hormone therapy on psychological functioning and quality of life in transgender individuals. Transgend Health 2016;1:21-31.
34. Nolan BJ, Zwickl S, Locke P, et al. Early access to testosterone therapy in transgender and gender-diverse adults seeking masculinization: a randomized clinical trial. JAMA Netw Open 2023;6:e2331919.
35. Buncamper ME, van der Sluis WB, van der Pas RSD, et al. Surgical outcome after penile inversion vaginoplasty: a retrospective study of 475 transgender women. Plast Reconstr Surg 2016;138:999-1007.
36. Hontscharuk R, Alba B, Hamidian Jahromi A, Schechter L. Penile inversion vaginoplasty outcomes: Complications and satisfaction. Andrology 2021;9:1732-43.
37. Papadopulos NA, Lellé JD, Zavlin D, et al. Quality of life and patient satisfaction following male-to-female sex reassignment surgery. J Sex Med 2017;14:721-30.
38. Klassen AF, Kaur M, Johnson N, et al. International phase I study protocol to develop a patient-reported outcome measure for adolescents and adults receiving gender-affirming treatments (the GENDER-Q). BMJ Open 2018;8:e025435.
39. van der Sluis WB, Schäfer T, Nijhuis THJ, Bouman MB. Genital gender-affirming surgery for transgender women. Best Pract Res Clin Obstet Gynaecol 2023;86:102297.
40. Tordoff DM, Wanta JW, Collin A, et al. Mental health outcomes in transgender and nonbinary youths receiving gender-affirming care. JAMA Netw Open 2022;5:e220978.
41. Shelemy L, Cotton S, Crane C, Knight M. Systematic review of prospective adult mental health outcomes following affirmative interventions for gender dysphoria. Int J Transgend Health 2024;26:480-500.
42. Bränström R, Pachankis JE. Reduction in mental health treatment utilization among transgender individuals after gender-affirming surgeries: a total population study. Am J Psychiatry 2020;177:727-34.
43. Campbell T, Mann S, Rodgers YVM, Tran NM. Mental health of transgender youth following gender identity milestones by level of family support. JAMA Pediatr 2024;178:870-8.
44. Boase ER, McLaren S. Discrimination and suicide risk among transgender and gender-diverse adults: The moderating roles of self-compassion, self-coldness, and gender identity. Psychol Sex Orientat Gend Divers 2024;11:606-18.
45. Dahlen S, Connolly D, Arif I, et al. International clinical practice guidelines for gender minority/trans people: systematic review and quality assessment. BMJ Open 2021;11:e048943.
46. Zwickl S, Wong AFQ, Dowers E, et al. Factors associated with suicide attempts among Australian transgender adults. BMC Psychiatry 2021;21:81.
47. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
48. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. J Med Libr Assoc 2021;109:174-200.
49. Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002.
50. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539-58.
51. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010;36:1-48.
52. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. Res Synth Methods 2011;2:61-76.
53. Baujat B, Mahé C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med 2002;21:2641-52.
54. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315: 629-34.
55. 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. J Adolesc Health 2022;70:643-9.
56. Hughto JMW, Gunn HA, Rood BA, Pantalone DW. Social and medical gender affirmation experiences are inversely associated with mental health problems in a U.S. non-probability sample of transgender adults. Arch Sex Behav 2020;49:2635-47.
57. Summers NA, Huynh TT, Dunn RC, et al. Effects of gender-affirming hormone therapy on progression along the HIV care continuum in transgender women. Open Forum Infect Dis 2021;8:ofab404.

58. Colizzi M, Costa R, Todarello O. Transsexual patients' psychiatric comorbidity and positive effect of cross-sex hormonal treatment on mental health: results from a longitudinal study. Psychoneuroendocrinology 2014;39:65-73.

59. Tucker RP, Testa RJ, Simpson TL, et al. Hormone therapy, gender affirmation surgery, and their association with recent suicidal ideation and depression symptoms in transgender veterans. Psychol Med 2018;48:2329-36.

Online supplementary material:

Supplementary Material 1. Search strategies, eligibility criteria, and record management details.

Supplementary Material 2. Reviewer roles at each review stage.

Supplementary Material 3. Study-level quality judgments.

Received: 25 March 2026; Accepted: 27 March 2026.

Contributions: all the authors made a substantial intellectual contribution, read and approved the final version of the manuscript, and agreed to be accountable for all aspects of the work.

Conflict of interest: the authors declare no conflict of interest regarding the publication of this article.

Ethics approval and consent to participate: not applicable.

Informed consent: not applicable.

Patient consent for publication: not applicable.

Availability of data and materials: data and materials, including supplementary materials, are available from the corresponding author upon request.