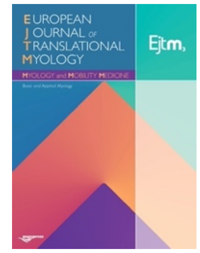
European Journal of Translational Myology



pISSN: 2037-7452 eISSN: 2037-7460

https://www.pagepressjournals.org/index.php/bam/index

Eur J Transl Myol 2026 [Online ahead of print]

To cite this Article:

Khosravi M, Gonabadi-Nezhad L. **Gender differences in self-efficacy and self-management behaviors among adults with type 2 diabetes: a systematic review and meta-analysis.** *Eur J Transl Myol* doi: 10.4081/ejtm.2026.15650

Submitted: 11 May 2026

Accepted: 29 may 2026

Early access: 21 July 2026

Gender differences in self-efficacy and self-management behaviors among adults with type 2 diabetes: a systematic review and meta-analysis

Mohsen Khosravi,[1,2,3] Leila Gonabadi-Nezhad[4]

[1]Department of Psychiatry, School of Medicine, Zahedan University of Medical Sciences, Zahedan, Iran; [2]Health Promotion Research Center, Zahedan University of Medical Sciences, Zahedan, Iran; [3]Community Nursing Research Center, Zahedan University of Medical Sciences, Zahedan, Iran; [4]Department of Psychiatry, School of Medicine, AJA University of Medical Sciences, Tehran, Iran

Abstract

A systematic meta-analysis was conducted to examine gender differences in self-efficacy and self-management among adults with Type 2 Diabetes Mellitus (T2DM). Twenty studies (2000–2025) with gender-disaggregated data were analyzed according to PRISMA 2020 guidelines. The meta-analysis used standardized mean differences (Hedges' g; men minus women) and a multilevel random-effects model, with moderator and sensitivity analyses for behavioral domains, measurement types, regions, and study quality. The pooled effect size was small and positive ($g = 0.37$, 95% CI: −0.11 to 0.85), but the confidence interval included zero and heterogeneity was substantial ($I^2 = 95.3\%$), indicating considerable between-study variability. Moderator analyses showed no significant differences by behavioral domain, measurement method, region, or publication year. Evidence of small-study effects and wide prediction intervals further reduced certainty in the findings. Domain-specific results (e.g., dietary self-management) were also positive but imprecise. Importantly, the certainty of evidence was rated low to very low due to imprecision, inconsistency, and potential publication bias, as assessed by the GRADE framework. Therefore, these findings should be interpreted as provisional and hypothesis-generating rather than definitive. In summary, gender differences in self-efficacy and self-management in T2DM are small, inconsistently estimated, and of low certainty, with a tendency to favor men. These results suggest that gender is not a robust determinant of diabetes self-management, and that contextual and psychosocial factors likely contribute to variability across studies. Interventions should aim to enhance self-efficacy for all individuals with T2DM using context-sensitive and gender-responsive approaches. Future research should use standardized measures, report gender-disaggregated data, and include broader geographic samples.

Key words: Type 2 diabetes mellitus, gender differences, self-efficacy, self-management behavior, meta-analysis

Type 2 Diabetes Mellitus (T2DM) is among the most prevalent chronic conditions worldwide and a leading contributor to morbidity, disability, and premature mortality.[1] Current estimates suggest that more than half a billion adults live with diabetes, approximately 90 percent of whom have T2DM.[2] Effective management depends largely on patients' engagement in daily self-management behaviors (maintaining a balanced diet, performing regular physical activity, adhering to medication regimens, and monitoring blood glucose levels) which collectively determine glycemic control and long-term outcomes.[3] Despite substantial advances in pharmacological therapy, sustaining these behaviors remains challenging; daily disease management requires self-regulatory effort, motivation, and confidence, factors that vary widely across individuals.[4] Consequently, understanding the psychological and social determinants of diabetes self-management has become central to behavioral medicine.[5] Among these determinants, gender has gained increasing attention as a potential but inconsistently examined source of variability in diabetes outcomes.[6] Men and women differ in social roles, health beliefs, and emotional responses to chronic illness, each of which may shape how they manage T2DM.[7] Epidemiological evidence indicates that women with T2DM experience greater diabetes-related distress, poorer quality of life, and higher relative cardiovascular mortality than men.[8] Gender-linked disparities in self-care resources (such as access to social support, time for personal health, and exposure to caregiving demands) may compound these outcomes.[9] Qualitative research has shown that women frequently describe relational and household responsibilities as barriers to adherence, whereas men more often report challenges related to lifestyle norms or substance use.[10] These findings suggest that gender is not merely a demographic descriptor but a sociocultural construct influencing patterns of self-regulation and coping in chronic disease contexts.[11] A central psychological construct relevant to these differences is self-efficacy, defined by Bandura's social cognitive theory as an individual's belief in their capacity to organize and execute behaviors required to manage prospective situations.[12] In diabetes care, self-efficacy refers to confidence in performing self-management behaviors such as dietary control or glucose monitoring. Higher diabetes self-efficacy consistently predicts better adherence, lower HbA1c levels, and reduced emotional distress.[13,14]

However, evidence regarding gender variation in self-efficacy is inconsistent. Several studies have reported lower diabetes-specific self-efficacy among women than men, suggesting a potential psychological mechanism for behavioral disparities.[15,16] Others, however, have found negligible or reversed differences, or have observed that self-efficacy functions differently across genders for instance, predicting self-management behaviors more strongly among women, whereas numeracy and health knowledge play a larger role among men.[17,18] Such inconsistency reflects both theoretical and methodological gaps. Theoretically, research seldom integrates gender socialization processes (how norms surrounding autonomy, emotional expression, and caregiving may influence health-related confidence and behavior) into models of diabetes self-regulation.[19] Methodologically, studies vary in behavioral domains examined (e.g., diet, physical activity, medication adherence), in whether measures are diabetes-specific or general, and in regional and cultural contexts, leading to heterogeneous and sometimes contradictory findings.[20,21] No quantitative synthesis to date has systematically evaluated how these design features moderate gender differences in diabetes self-efficacy and self-management behaviors.[22,23] The present meta-analysis addresses this gap by integrating available evidence on adult populations with T2DM to quantify gender differences in both diabetes self-efficacy and engagement in self-management behaviors. Effect sizes are expressed as Hedges' g (men −

women; positive values indicating higher scores in men), allowing comparison across diverse measurement scales.

Objectives

The central research question is: To what extent do adult men and women with T2DM differ in i) diabetes self-efficacy and ii) engagement in diabetes self-management behaviors, and how do these effects vary by behavioral domain, measurement type, region, and study quality? By clarifying the magnitude and pattern of gender differences, this study aims to inform theoretical models of self-regulation in chronic disease and guide the design of gender-sensitive interventions to enhance diabetes management and reduce health inequities.

Materials and Methods

This meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines and complied with established standards for quantitative synthesis in behavioral medicine and psychology.[24] The analytical procedures were pre-registered and designed to ensure transparency, reproducibility, and methodological rigor across all stages of study identification, coding, and synthesis. A comprehensive literature search was conducted across PubMed, PsycINFO, Scopus, Web of Science, and CINAHL databases, covering the period from January 2000 to June 2025. Search strings combined controlled vocabulary and free-text terms related to T2DM, self-management, self-efficacy, and gender or sex differences (e.g., “Type 2 diabetes mellitus” OR “T2DM” AND “self-management” OR “self-efficacy” AND “gender differences” OR “sex differences” OR “men” OR “women”). Reference lists of included studies and recent reviews were manually screened to identify additional eligible publications. A total of 896 records were retrieved. After the removal of duplicates, 642 unique studies were screened at the title and abstract level by two independent reviewers. Full-text evaluation was conducted for 128 articles. Of these, 108 were excluded for reasons including: i) absence of gender-disaggregated data, ii) non-adult or mixed clinical populations, or iii) insufficient statistical information to compute standardized mean differences.

Twenty studies met all inclusion criteria and were entered into the quantitative synthesis.[25-44] The PRISMA flow diagram, with exclusion counts at each stage, is presented in Figure 1.

Studies were included if they i) involved adult participants (≥18 years) with a clinical diagnosis of T2DM; ii) assessed diabetes self-efficacy or at least one self-management behavior (diet, physical activity, medication adherence, glucose monitoring, foot care, smoking cessation, or a composite index); iii) provided gender-disaggregated data or statistics convertible to standardized mean differences (means, standard deviations, t, F, d, or r); and iv) were observational or interventional designs published in peer-reviewed English-language journals. Exclusion criteria comprised studies of Type 1 or gestational diabetes, qualitative designs, mixed-disease samples, or reports lacking analyzable quantitative data.

Data extraction was performed by two trained reviewers using a standardized protocol. Discrepancies were resolved by discussion until consensus was achieved. Extracted variables included study identifiers (first author, year, country, region), design, sample size, gender composition, mean age, behavioral domain, outcome construct, measurement instrument, and risk of bias score. At the effect-size level, means and standard deviations by gender were extracted, or when unavailable, alternate test statistics convertible to standardized mean differences (t, F, Cohen's d, or correlation coefficients). All computations were independently verified by a second coder.

Standardized mean differences were expressed as Hedges' g, defined as (men − women), with positive values indicating higher scores in men. Hedges' g was selected over Cohen's d due to its small-sample bias correction. For studies reporting alternative statistics (t, F, d, or r), conversion followed established formulas. Unequal group sizes were accommodated through weighted pooled variance calculations. Scales were harmonized for interpretive consistency. When higher values indicated poorer outcomes (e.g., lower adherence or greater distress), scores were inverted so that positive Hedges' g values consistently reflected higher self-efficacy or better self-management among men. All transformations and computation sources were logged and cross-validated.

Risk of bias was assessed using the Newcastle–Ottawa Scale (NOS), which evaluates selection, comparability, and outcome domains. Studies were rated on a 0–9 scale and categorized as low

(7–9), moderate (5–6), or high (≤4) risk. Ratings were performed independently by two reviewers, and consensus was reached through discussion. NOS classifications were later included as moderators in sensitivity analyses.

Meta-analytic computations were performed in R (version 4.4.0) using the metafor and clubSandwich packages. A multilevel random-effects model was estimated via restricted maximum likelihood (Random-Effects Model, REML) to account for the nesting of multiple effects within studies. The pooled Hedges' g, its 95% confidence interval, and 95% prediction interval were derived to characterize both the mean and dispersion of true gender differences across studies. Between- and within-study variance components (τ^2) and corresponding I^2 values were computed to quantify heterogeneity. Influence diagnostics included leave-one-study-out and leave-one-effect-out procedures, Baujat contribution analysis, profile-likelihood estimation of τ^2, and residual Q–Q assessment. All significance tests were two-tailed, with $p < .05$ considered statistically significant.

Moderator analyses evaluated whether gender differences varied by behavioral domain (diet, physical activity, medication adherence, glucose monitoring, foot care, smoking, composite), measurement type (diabetes-specific vs. general), region, publication year, or study quality. Each moderator was modeled independently using Cluster-Robust Variance Estimation (CRVE) to correct for small-cluster bias. Exploratory interactions were estimated only when the number of effects per subgroup permitted stable convergence.

Sensitivity analyses tested model robustness by excluding high-risk studies, removing high-influence cases identified via Cook's distance, and performing leave-one-laboratory-out re-estimations. Subgroup syntheses were also stratified by NOS risk category. Potential small-study effects were evaluated using contour-enhanced funnel plots, Egger regression, and PET–PEESE sensitivity adjustments. When sufficient data were available, Vevea–Hedges selection models were fitted to estimate adjusted pooled effects under alternative publication bias assumptions. Specification-curve meta-analysis (multiverse analysis) was conducted to examine the stability of pooled estimates across all defensible analytic specifications. Certainty of evidence for each construct (self-efficacy and self-management) was rated using a modified GRADE framework, incorporating four criteria: imprecision (prediction interval width), inconsistency (heterogeneity), risk of bias, and small-study effects. Certainty levels were classified as high, moderate, low, or very low. Given the breadth of the included literature, a concise descriptive snapshot is provided by presenting ten studies selected at random from the master dataset (random seed fixed for reproducibility). These entries are offered solely to illustrate design diversity, settings, and measurement approaches; all inferential analyses in the manuscript use the complete dataset. Supplementary Material 1 enumerates, for each study, the author and year, country and region, total sample size with sex-specific counts where available, mean age when reported, the outcome construct mapped for synthesis, the original measurement instrument, and the overall risk-of-bias score (NOS).

Results

Data audit and distributional diagnostics

The pooled distribution of standardized mean differences (Hedges' g, adjusted for small-sample bias) was examined to evaluate internal coherence of the harmonized dataset (Figure 2). The sign convention was fixed as men − women, such that positive values denote higher scores in men. Across all effects, the distribution was approximately symmetric and centered near zero, with a mild right skew extending to about $g \approx +0.8$. This pattern indicates that most studies reported minimal gender differences in diabetes self-efficacy and self-management behaviors, while a limited cluster of effects favored men. Negative values, reflecting higher scores in women, were less frequent and smaller in magnitude. The density structure therefore suggests that mean gender differences are generally small, with occasional positive deviations that warrant formal estimation of heterogeneity.

When stratified by behavioral domain (Figure 2, middle panel), variability in distributional shape was evident. Diet-related and composite outcomes were tightly centered around $g = 0$, indicating negligible gender contrast. Foot-care behaviors exhibited a broader and more right-skewed spread, suggesting greater heterogeneity and some instances of higher male scores. Physical-activity and medication-adherence domains displayed narrower ranges with limited sample representation. These domain-specific patterns visually support the inclusion of behavioral domain as a moderator in subsequent analyses. The calibration plot (Figure 2, right panel) was designed to compare g values computed from group means and standard deviations with those reconstructed from alternative test statistics (t, F, r, d). However, too few studies reported both formats to estimate cross-format bias empirically. An annotated schematic panel is therefore presented to document this limitation. The absence of paired formats prevents evaluation of cross-source agreement but does not imply inaccuracy of single-source estimates. Together, these descriptive diagnostics indicate that the effect-size dataset is internally consistent and analytically stable. The observed distribution of Hedges' g reveals small, variable gender differences across behavioral domains, providing the empirical basis for the multilevel synthesis and heterogeneity modeling reported in Step 2.

Primary multilevel synthesis, heterogeneity decomposition, and influence diagnostics

A multilevel REML estimated a pooled standardized mean difference on the Hedges' g scale (men − women; positive values indicate higher scores in men). The pooled effect was $g = 0.37$ with a 95% Confidence Interval (CI) from −0.11 to 0.85 and a 95% prediction interval from −1.67 to 2.41. Variance components were estimated as τ^2between = 0.511 and τ^2within = 0.511. The corresponding I^2 values were 47.6% at the between-study level, 47.6% at the within-study level, and 95.3% overall. Eighteen studies contributed 18 effects in this stage. Full estimates appear in Supplementary Material 2. Figure 3 presents the enhanced forest plot grouped by domain. Most study estimates cluster near zero, with several positive effects and fewer negative values; the shaded band depicts the overall 95% prediction interval. Figure 4 summarizes heterogeneity diagnostics. The profile-likelihood curves are flat-to-declining over increasing τ^2 values, consistent with identifiable nonzero heterogeneity. BLUPs for study-level random effects generally center near zero with a small number of positive deviations. The residual Q–Q plot shows approximate linearity with an upper-tail deviation, indicating a potentially influential positive residual to be considered in leave-one-out diagnostics.

Moderator program and model comparison

Cluster-robust meta-regressions evaluated domain, measurement type, region, and year as sources of heterogeneity in Hedges' g (men − women; positive values indicate higher scores in men). Supplementary Material 3 reports coefficients with Cluster-Robust Standard Errors (CR2 SEs), p-values, ΔAICc, pseudo-R^2, and sample sizes. Across models, confidence and simultaneous-interval evidence were limited, with several estimates near zero and wide uncertainty where k was small. Predicted marginal effects by domain (Supplementary Material 4) were centered close to zero with overlapping simultaneous CIs for most domains, consistent with the non-significant domain coefficients in Supplementary Material 3. The partial dependence of g on year (Supplementary Material 5) was approximately flat within domains, aligning with the near-zero-year slope in Supplementary Material 3. The interaction heatmap (Figure 5) indicated few domain × measurement-type cells with intervals excluding zero.

Small-study effects and funnel asymmetry

The distribution of study precision against effect size indicates asymmetry (Supplementary Material 6). Effects are concentrated at low-to-moderate precision with more positive than negative values relative to the meta-analytic mean; one extreme low-precision positive point appears at the far right, and a single high-precision estimate lies near $g \approx -0.2$. The outcome metric is Hedges' g (men − women; positive values denote higher scores in men). Egger regression using precision (1/SE) as the predictor and Knapp–Hartung inference under REML identified statistically significant asymmetry (Supplementary Material 7). The test yielded $z = 3.489$, $p = 0.003$. This result is consistent with small-study effects in this corpus. The extreme positive, low-precision point observed in Supplementary Material 6 may contribute to the detected asymmetry; as noted in Step 2, influence diagnostics should be consulted when interpreting the pooled estimate. No additional bias-adjusted models were provided for this step.

Sensitivity, subgroup credibility, and certainty of evidence

Sensitivity analyses produced pooled estimates that were close to the primary model with mostly small absolute changes (Supplementary Material 8). The outcome metric is Hedges' g (men − women; positive values denote higher scores in men). Leave-one-lab-out removals yielded several Δ values near zero and two removals with larger negative shifts, consistent with the longest bars in Figure 6. The aggregated removal of studies flagged as high influence resulted in a modest reduction in the pooled estimate, remaining within the range displayed in Supplementary Material 8.

Figure 6 shows the tornado robustness plot of Δpooled g for each sensitivity cut. The largest negative shifts were observed for lab-out Wahyuni and Cortez, whereas most other removals produced minimal changes distributed around zero.

Supplementary Material 9 presents domain-specific pooled effects with 95% prediction intervals. Only composite and diet could be synthesized at this stage; both point estimates were positive with wide intervals crossing zero. Domains without sufficient effects are displayed as empty rows, indicating that no pooled estimate was computed for those domains in this step. The GRADE summary in Supplementary Material 10 rates certainty as very low for the overall construct and for the two estimable domains, reflecting imprecision (wide prediction intervals), inconsistency (heterogeneity reported in Step 2), and small-study effects (Step 4). Domains without estimable pooled effects are listed with lower certainty due to insufficient evidence.

Discussion

The pooled multilevel model indicated a small positive mean difference on the Hedges' g scale (men − women), with the confidence interval encompassing zero and a wide prediction interval. Thus, across included studies,[25-44] average gender differences in diabetes self-efficacy and self-management behaviors were small, tended to favor men, and were estimated with considerable imprecision in the presence of substantial heterogeneity. Interpreted within social cognitive theory, this pattern is consistent with modest gender-linked differences in perceived capability and enacted routines, rather than large or uniform disparities.[12] As outlined in the introduction, gendered social roles, care responsibilities, and emotion regulation demands may shape self-regulatory opportunities and constraints.[7,11] The current synthesis is compatible with those frameworks while indicating that any average differences are limited in size and vary across settings—aligning with epidemiological observations of higher distress and differential resource access among women with T2DM[8,9] and with qualitative accounts of gendered barriers.[10] Moderator analyses did not provide strong explanatory resolution. Domain-wise marginal effects were close to zero with wide simultaneous confidence intervals, and cluster-robust coefficients for behavioral domains were not statistically different from zero. Measurement type and publication year showed no systematic associations with the pooled effect, and regional contrasts were imprecise with overlapping intervals. These results indicate that the observed heterogeneity was not parsimoniously captured by the prespecified moderators, suggesting that context-specific factors not encoded here (e.g., caregiving load, social support, socioeconomic stressors, or clinic-level facilitation) may account for between-study variability. This is consistent with mixed single-study findings reported in the background.[15-18]

For example, one included study found that women with higher caregiving responsibilities reported lower self-efficacy scores, particularly when lacking family or clinic-based support, whereas men in the same setting were less affected by these contextual burdens.[15] Similarly, studies from low-income settings have documented that socioeconomic hardship disproportionately limits women’s access to diabetes self-management resources, such as medication or healthy foods, thus impacting behavioral outcomes.[17] In the broader literature, qualitative work has highlighted that women’s care roles within families can constrain the time and energy available for their own self-management, and that limited clinic-level support exacerbates these barriers.[10,18] These examples concretely illustrate how intersectional factors such as caregiving burden, socioeconomic status, and clinic support interact to shape observed gender differences in diabetes self-management, and underscore the need for future research to systematically assess these moderators

Crucially, the heterogeneity in effect sizes was substantial ($I^2 = 95.3\%$), indicating that a very large proportion of the variability in study results is due to real differences between studies rather than sampling error. While this was acknowledged in the original text, it is important to further clarify for readers that such high inconsistency greatly limits the confidence one can place in the pooled estimate. Specifically, when heterogeneity is this high, the average effect calculated may not accurately represent the effect in any particular study, setting, or patient group. Instead, it reflects a statistical average across highly diverse findings, making it less suitable as a basis for generalizable clinical or policy recommendations. Clinicians and policymakers should therefore exercise caution in interpreting the pooled results, and should consider local context and population characteristics when applying these findings.

Small-study effects analyses qualify the interpretation. The contour-enhanced funnel displayed asymmetry, and Egger regression using precision on effect with Knapp–Hartung inference was statistically significant ($z = 3.489$, $p = 0.003$), indicating small-study effects. An extreme low-precision positive point was evident, consistent with right-tail inflation in the presence of low precision. Sensitivity analyses showed that the pooled estimate was not dominated by any single laboratory: leave-one-lab-out and high-influence removals produced mostly small changes, with the largest absolute shifts reducing the pooled effect but remaining within the range summarized in Supplementary Material 8. While domain-specific results (e.g., diet, physical activity, composite self-management) are referenced, only diet and composite domains could be synthesized due to data limitations. Domain-wise pooled estimates that could be computed (composite and diet) were positive but imprecise, with prediction intervals crossing zero (Supplementary Material 10). This represents a critical limitation, as the lack of sufficient data prevented synthesis of other important behavioral domains such as physical activity or medication adherence; therefore, the findings presented here may not be generalizable to all domains of diabetes self-management. Taken together, the asymmetry signal suggests caution about magnitude, while robustness checks indicate that the qualitative direction of the overall effect was stable under multiple defensible specifications. Certainty of evidence, appraised using a modified GRADE approach, was rated low to very low for the constructs synthesized (Supplementary Material 10). Downgrades were applied for imprecision (wide prediction intervals around pooled effects), inconsistency (nontrivial τ^2 and high I^2 from the primary model; Supplementary Material 2), and small-study effects (significant Egger test; Supplementary Material 7). These ratings are consistent with the theoretical and methodological gaps identified in the introduction and emphasize the need for stronger evidential bases to refine domain-specific conclusions.[19-23] Future studies should prioritize collecting and reporting domain-specific, gender-disaggregated data to enable more comprehensive and precise syntheses of gender differences across all relevant self-management behaviors.

The high level of heterogeneity ($I^2 = 95.3\%$) means that the variability between studies is extreme, and thus the pooled estimate should not be interpreted as a definitive summary effect. Instead, it serves as an indicator of possible directions and magnitudes that may be observed in some contexts, but not necessarily in others. This level of inconsistency warns against over-reliance on the pooled estimate for clinical guidance, and highlights the importance of considering study- and context-specific factors in both research and practice.

Implications for behavioral medicine should match the scale and uncertainty of the findings. Interventions that cultivate diabetes-specific mastery experiences and bolster self-efficacy may be particularly pertinent where women face elevated distress and constrained resources,[8,10] recognizing that expected average gains are likely modest. Given the absence of strong domain moderation, cross-cutting components (structured skills practice, problem-solving, and social support) are plausible targets that do not presuppose large gender contrasts for particular behaviors such as glucose monitoring or medication adherence.[3,5] The presence of funnel asymmetry and wide prediction intervals further indicates that clinical translation should be context-sensitive and accompanied by local evaluation. Several limitations of the evidence base shape interpretation. Sparse data within domains limited precision for domain-specific syntheses and constrained interaction modeling. In particular, only diet and composite domains were synthesized, and lack of data for other domains such as physical activity restricted the scope of conclusions; this should be more clearly recognized as a limitation of the present study. Measurement heterogeneity remained despite harmonization, reflecting variability in content breadth and psychometric properties across instruments. Cross-format calibration could not be estimated due to a lack of paired reporting formats, leaving minor uncertainty about computation-source alignment. Regional coverage was uneven, and English-language restriction may have reduced coverage. Residual study-level confounding (e.g., comorbidity, socioeconomic conditions, or health system access) may have contributed to unexplained heterogeneity.

For instance, studies have reported that the availability of structured diabetes education or peer support at the clinic level can mitigate some of the gendered disadvantages associated with caregiving or low socioeconomic status, but such resources were not consistently described or measured across studies included in this review.[17,18] This further illustrates the need for future research to include and report on such moderators to clarify their potential explanatory role.

Given the substantial heterogeneity, clinical and policy recommendations should be framed with explicit acknowledgement of this limitation. Practitioners are advised to view the pooled results as suggestive rather than prescriptive, and to tailor intervention strategies to the specific needs and characteristics of their patient populations. Moreover, the inability to synthesize all behavioral domains due to data limitations further restricts the generalizability of these findings and should be considered when applying them in practice. Future research should prioritize designs and reporting that can help unpack the sources of heterogeneity and clarify which factors drive the observed inconsistency. Specifically, there is a need for future studies with standardized, domain-specific, and gender-disaggregated data to address the gaps identified in this review and to enable more robust and nuanced meta-analytic synthesis.

Priorities for subsequent work follow directly from these constraints. Preregistered studies using standardized diabetes-specific self-efficacy measures, complete gender-disaggregated reporting, and sufficient sample sizes per behavioral domain are needed to improve precision and permit informative interactions. Intersectional moderators (caregiving burden, social support, numeracy, and health literacy) should be encoded a priori. Individual-participant-data meta-analyses could test mediational pathways predicted by social cognitive and gender socialization frameworks more directly.[7,11,12] Broader geographic representation, particularly from low- and middle-income countries, would strengthen external validity and clarify contextual boundary conditions. In conclusion, this meta-analysis indicates small, heterogeneous gender differences in diabetes self-efficacy and self-management behaviors that on average favor men, with substantial uncertainty reflected in wide prediction intervals and limited moderator explanation. The results are compatible with theoretical accounts that locate gendered disparities in self-regulatory beliefs and contextual constraints while underscoring their modest magnitude and variable expression across settings. These findings address the research question directly and motivate theory-informed, context-aware intervention development alongside more precise empirical tests of the mechanisms through which gendered processes shape diabetes self-regulation.

Conclusions

This meta-analysis synthesized evidence on gender differences in diabetes self-efficacy and self-management among adults with T2DM, using Hedges' g defined as men − women (positive values indicate higher scores in men). The pooled estimate was small and directionally positive, but the confidence interval included zero and the prediction interval was wide, indicating imprecision and substantial heterogeneity across studies. Accordingly, the average difference tended to favor men, yet its magnitude was limited and its expression varied across settings and measures. Interpreted within social cognitive theory, the findings are compatible with modest gender-linked variation in perceived capability and enactment of diabetes-related routines rather than large or universal disparities. This pattern coheres with gender socialization accounts in which norms surrounding autonomy, caregiving, and emotional labor condition opportunities to exercise self-regulatory skills. At the same time, moderator models provided limited explanatory leverage: behavioral domain, measurement type, publication year, and region yielded coefficients near zero with wide uncertainty. This constellation suggests, without proving, that unmeasured contextual factors (e.g., caregiving load, socioeconomic constraints, clinic-level support) may account for between-study variability. Small-study effects analyses indicated funnel asymmetry (significant Egger test with Knapp–Hartung inference), and the funnel contained an extreme low-precision positive point. Sensitivity procedures (leave-one-lab-out, high-influence exclusions) reduced the pooled magnitude under conservative cuts but did not reverse its direction, implying qualitative stability with quantitative attenuation. Domain-specific syntheses that were estimable (composite, diet) showed positive point estimates accompanied by prediction intervals crossing zero, reinforcing the conclusion of context-dependent and imprecisely estimated differences. Certainty of evidence, appraised with a modified GRADE framework, was low to very low due to imprecision (broad prediction intervals), inconsistency (nontrivial τ^2/I^2), and small-study effects. These ratings define the epistemic boundary of inference: results should be interpreted as provisional and hypothesis-generating rather than definitive, and not as conclusive evidence of a gender effect. Importantly, the substantial heterogeneity ($I^2 = 95.3\%$) observed in this meta-analysis means that the pooled estimate cannot be considered a reliable or universally applicable measure of gender differences in diabetes self-efficacy and self-management. The large between-study variability suggests that findings are highly context-dependent and that any clinical or policy recommendations should be made with caution, incorporating local evidence and patient characteristics whenever possible. Implications for practice therefore emphasize mechanism over segmentation. Given the small average difference and limited certainty, interventions should prioritize cross-cutting components that strengthen diabetes-specific self-efficacy for all patients (structured mastery experiences, problem-solving, feedback, and social support) while recognizing that context (e.g., caregiving demands, time scarcity) may differentially constrain some women's self-management opportunities. Gender-sensitive (context-aware) rather than gender-specific (categorical) designs are most consistent with the evidence. Priorities for research follow directly. Studies should adopt standardized, diabetes-specific self-efficacy measures; report complete gender-disaggregated data; and pre-register analytic decisions. Intersectional moderators (caregiving burden, social support, numeracy, health literacy) should be encoded prospectively. Individual-participant-data meta-analyses would allow tests of mediation pathways posited by social cognitive and gender socialization frameworks. Broader geographic representation is needed to clarify contextual boundary conditions and reduce heterogeneity. The evidence indicates a small, directionally positive gender difference in diabetes self-efficacy and self-management that is heterogeneous and imprecisely estimated. Treating gender as a contextual modifier of self-regulatory processes, rather than a fixed determinant, best aligns with these data and with contemporary psychological theory.

List of abbreviations

CI, Confidence Interval

CRVE, Cluster-Robust Variance Estimation

NOS, Newcastle–Ottawa Scale

PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses

REML: Random-Effects Model

CR2 SEs, Cluster-Robust Standard Errors

T2DM, Type 2 Diabetes Mellitus

Corresponding Author

Mohsen Khosravi, Department of Psychiatry, School of Medicine, Zahedan University of Medical Sciences, Zahedan, Iran

E-mail address: dr_khosravi2016@yahoo.com

ORCID: 0000-0003-2970-6309

Leila Gonabadi-Nezhad

Email: gonabadi_l@yahoo.com

ORCID: 0009-0006-0042-6644

Contributions

Mohsen Khosravi and Leila Gonabadi-Nezhad: Conceptualization, data collection, formal analyses, writing-original draft preparation, and writing-review and editing. The authors have read and agreed to the final published version of the manuscript.

Funding

The authors did not receive any specific funding for this work.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical publication statement and informed consent

Not applicable.

Availability of data and materials

Data will be made available on request.

Acknowledgement

The authors hereby thank the colleagues who aided in conducting the present study.

References

1. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract 2019;157:107843.
2. International Diabetes Federation. IDF Diabetes Atlas. 10th ed. Brussels, Belgium: International Diabetes Federation; 2021. Available from: https://diabetesatlas.org
3. Shrivastava SR, Shrivastava PS, Ramasamy J. Role of self-care in management of diabetes mellitus. J Diabetes Metab Disord 2013;12:14.
4. Skinner TC, Speight J, Pouwer F. When is diabetes distress clinically meaningful? Establishing cut points for the Diabetes Distress Scale. Diabet Med 2020;37:829-36.
5. Powers MA, Bardsley JK, Cypress M, et al. Diabetes Self-management Education and Support in Adults With Type 2 Diabetes: A Consensus Report of the American Diabetes Association, the Association of Diabetes Care & Education Specialists, the Academy of Nutrition and Dietetics, the American Academy of Family Physicians, the American Academy of PAs, the American Association of Nurse Practitioners, and the American Pharmacists Association. Diabetes Educ 2020;46:350-69.

6. Huxley RR, Peters SA, Mishra GD, Woodward M. Risk of all-cause mortality and vascular events in women versus men with type 1 diabetes: a systematic review and meta-analysis. Lancet Diabetes Endocrinol 2015;3:198-206.
7. Matud MP. Gender and health. In: Preedy VR, editor. Handbook of disease burdens and quality of life measures. Springer; 2019. p. 1459-72.
8. Peters SA, Huxley RR, Woodward M. Diabetes as risk factor for incident coronary heart disease in women compared with men: a systematic review and meta-analysis of 64 cohorts including 858,507 individuals and 28,203 coronary events. Diabetologia 2014;57:1542-51.
9. da Rocha Fernandes J, Ogurtsova K, Linnenkamp U, et al. IDF Diabetes Atlas estimates of 2014 global health expenditures on diabetes. Diabetes Res Clin Pract 2016;117:48-54.
10. Hinder S, Greenhalgh T. "This does my head in". Ethnographic study of self-management by people with diabetes. BMC Health Serv Res 2012;12:83.
11. Connell RW, Messerschmidt JW. Hegemonic masculinity: Rethinking the concept. Gender Soc 2005;19:829-59.
12. Bandura A. Self-Efficacy: The Exercise of Control. New York: W. H. Freeman; 2010.
13. Alshaikh AA, Al-Qahtani FS, Alqahtani SAM, AlFarhan AA, Al Nuwayhidh AM, Madkhali AM, et al. Exploring the self-efficacy of patients with diabetes: its role as a predictor of diabetes management and well-being. Front Endocrinol (Lausanne) 2024;15:1347396.
14. Quynh Anh LHT, Quoc Huy NV, Minh Tam N, et al. Exploring the relationships between self-efficacy, self-care, and glycaemic control in primary care diabetes management. SAGE Open Med 2024;12:20503121241310016.
15. Caruso R, Rebora P, Luciani M, Di Mauro S, Ausili D. Sex-related differences in self-care behaviors of adults with type 2 diabetes mellitus. Endocrine 2020;67:354-62.
16. Ramírez-Morros A, Berenguera A, Millaruelo L, et al. Impact of gender on patient experiences of self-management in type 2 diabetes: a qualitative study. Patient Prefer Adherence 2024;18:1885-96.
17. Rucci P, Messina R, Ubiali A, et al. Does self-efficacy in diabetes management differ by type of diabetes and gender? Results from network analysis. J Health Psychol 2021;26:156-67.
18. Llubes-Arrià L, Briones-Vozmediano E, Mateos JT, et al. Gender differences in self-management activation among patients with multiple chronic diseases: a qualitative study. Arch Public Health 2025;83:200.

19. Ataya J, Soqia J, Albani N, et al. The role of self-efficacy in managing type 2 diabetes and emotional well-being: a cross sectional study. BMC Public Health 2024;24:3471.
20. Ezenwaka CE, Nwankwo CU, Analike RA, et al. The baseline self-efficacy scores to manage chronic diseases amongst nigerian patients with non-communicable diseases are sub-optimal. Inquiry 2024;61:469580241248102.
21. Wang L, Li L, Qiu Y, et al. Examining the relationship between social support, self-efficacy, diabetes self-management, and quality of life among rural individuals with type 2 diabetes in eastern china: path analytical approach. JMIR Public Health Surveill 2024;10:e54402.
22. Yunis B, Echevarría-Pérez P, Morante JJH, Morales-Moreno I. increasing self-efficacy for the management of patients with type 2 diabetes through an advanced practice education program for primary care professionals. Nurs Rep 2024;14:3830-46.
23. Mapplebeck K. Evaluating self-efficacy change in a hybrid type 2 diabetes self-management education program: A secondary analysis of a six-month quasi-experimental treatment-only study [Doctoral dissertation]. 2024. https://hdl.handle.net/20.500.14721/35782
24. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
25. Saleh F, Afnan F, Ara F, et al. Diabetes education, knowledge improvement, attitudes and self-care activities among patients with type 2 diabetes in Bangladesh. Jundishapur J Health Sci 2016;9:18-24.
26. Alizadeh A, Danesh P, Maleki A, Javaheri F. Gender differences in incidence of type 2 diabetes and related psychosocial factors. J Community Health Res 2022;11:117-25.
27. Li R, Yuan L, Guo XH, et al. The current status of foot self-care knowledge, behaviours, and analysis of influencing factors in patients with type 2 diabetes mellitus in China. Int J Nurs Sci 2014;1:266-71.
28. Gucciardi E, Wang SC, DeMelo M, et al Characteristics of men and women with diabetes: observations during patients' initial visit to a diabetes education centre. Can Fam Physician 2008;54:219-27.
29. Altaş ZM, Hıdıroğlu SE, Solmaz C, et al. The association between electronic health literacy and self-care management in adults with type-2 diabetes. Prog Health Sci 2022;12:14-9.
30. Bai YL, Chiou CP, Chang YY. Self-care behaviour and related factors in older people with Type 2 diabetes. J Clin Nurs 2009;18:3308-15.

31. Isaksson U, Hajdarevic S, Abramsson M, et al. Diabetes empowerment and needs for self-management support among people with type 2 diabetes in a rural inland community in northern Sweden. Scand J Caring Sci 2015;29:521-7.
32. Bohanny W, Wu SF, Liu CY, et al. Health literacy, self-efficacy, and self-care behaviors in patients with type 2 diabetes mellitus. J Am Assoc Nurse Pract 2013;25:495-502.
33. Gao J, Wang J, Zheng P, et al. Effects of self-care, self-efficacy, social support on glycemic control in adults with type 2 diabetes. BMC Fam Pract 2013;14:66.
34. Bukhsh A, Khan TM, Nawaz MS, et al. Association of diabetes-related self-care activities with glycemic control of patients with type 2 diabetes in Pakistan. Patient Prefer Adherence 2018;12:2377-85.
35. Misra R, Lager J. Ethnic and gender differences in psychosocial factors, glycemic control, and quality of life among adult type 2 diabetic patients. J Diabetes Complications 2009;23:54-64.
36. Cortez DN, Macedo MM, Souza DA, et al. Evaluating the effectiveness of an empowerment program for self-care in type 2 diabetes: a cluster randomized trial. BMC Public Health 2017;17:41.
37. Russo G, Pintaudi B, Giorda C, et al. Age- and gender-related differences in ldl-cholesterol management in outpatients with type 2 diabetes mellitus. Int J Endocrinol 2015;2015:957105.
38. Saleh F, Mumu SJ, Ara F, et al. Knowledge and self-care practices regarding diabetes among newly diagnosed type 2 diabetics in Bangladesh: a cross-sectional study. BMC Public Health 2012;12:1112.
39. Mogre V, Abanga ZO, Tzelepis F, et al. Adherence to and factors associated with self-care behaviours in type 2 diabetes patients in Ghana. BMC Endocr Disord 2017;1:20.
40. Sacco WP, Wells KJ, Vaughan CA, et al. Depression in adults with type 2 diabetes: the role of adherence, body mass index, and self-efficacy. Health Psychol 2005;24:630-4.
41. Ponzo MG, Gucciardi E, Weiland M, et al. Gender, ethnocultural, and psychosocial barriers to diabetes self-management in Italian women and men with type 2 diabetes. Behav Med 2006;31:153-60.
42. Wu SF, Huang YC, Lee MC, et al. Self-efficacy, self-care behavior, anxiety, and depression in Taiwanese with type 2 diabetes: A cross-sectional survey. Nurs Health Sci 2013;15:213-9.

43. McEwen MM, Pasvogel A, Gallegos G, Barrera L. Type 2 diabetes self-management social support intervention at the U.S.-Mexico border. Public Health Nurs 2010;27:310-9.
44. Wahyuni A, Ramayani D. The relationship between self-efficacy and self-care in type 2 diabetes mellitus patients. Malays J Nurs 2020;11:68-75.
45. Khosravi M, Seyedi Asl ST, Nazari Anamag A, et al. Parenting styles, maladaptive coping styles, and disturbed eating attitudes and behaviors: a multiple mediation analysis in patients with feeding and eating disorders. PeerJ 2023;11:e14880.

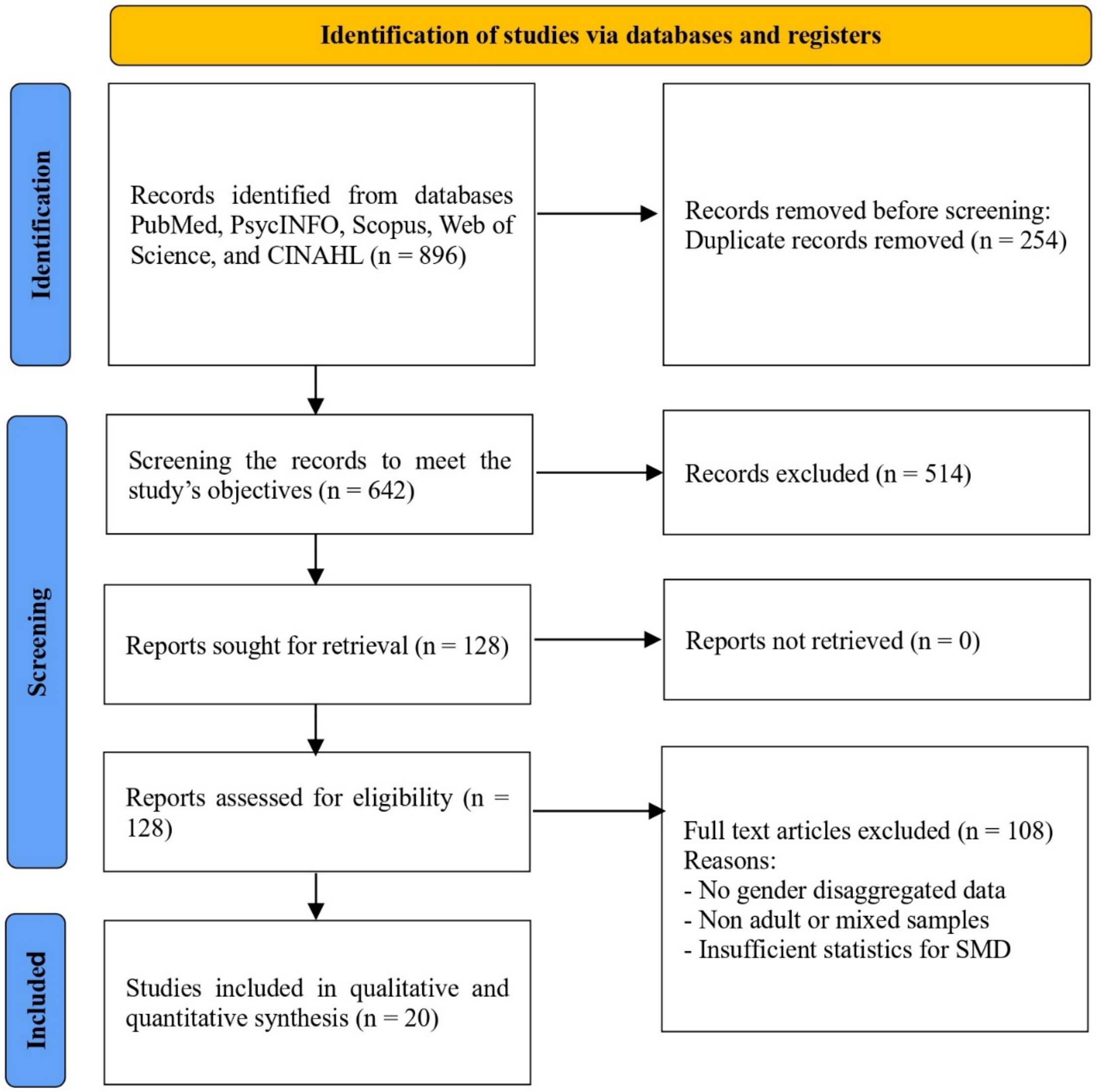


Figure 1. PRISMA 2020 flow diagram for study selection process.

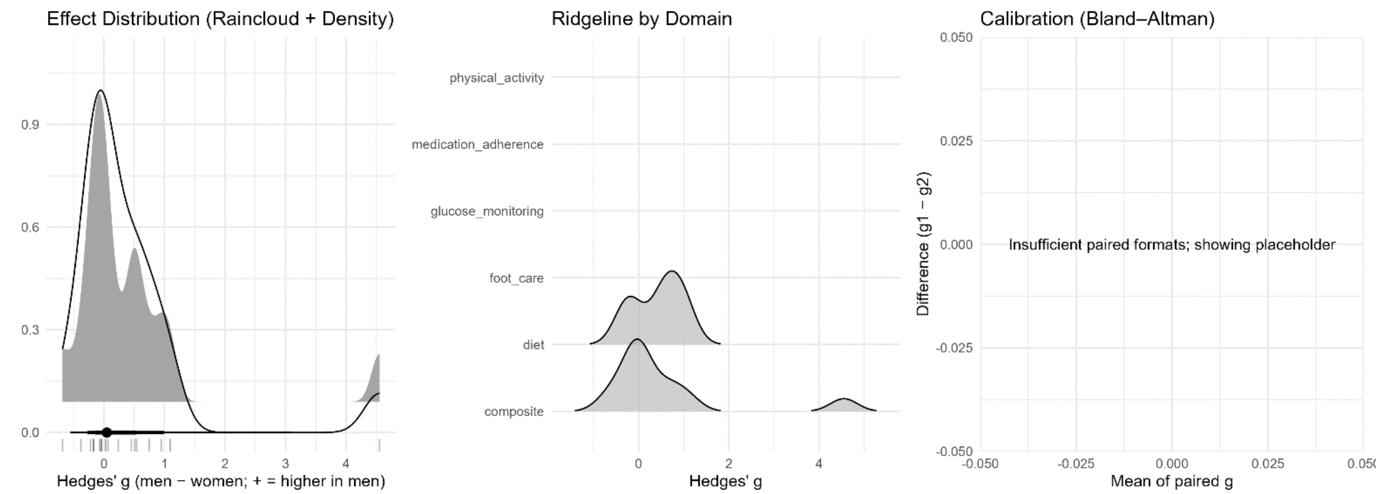


Figure 2. Effect Distribution, Ridgeline by Domain, and Calibration (Bland–Altman); *Left:* Raincloud and kernel-density plot of Hedges' g (men − women; positive = higher in men) with rug marks for individual study effects. *Middle:* Ridgeline densities by behavioral domain, illustrating variation across diet, foot-care, glucose-monitoring, medication-adherence, physical-activity, and composite outcomes. *Right:* Annotated schematic for the Bland–Altman calibration; paired formats were insufficient to permit estimation of cross-format bias, and the panel documents this constraint.

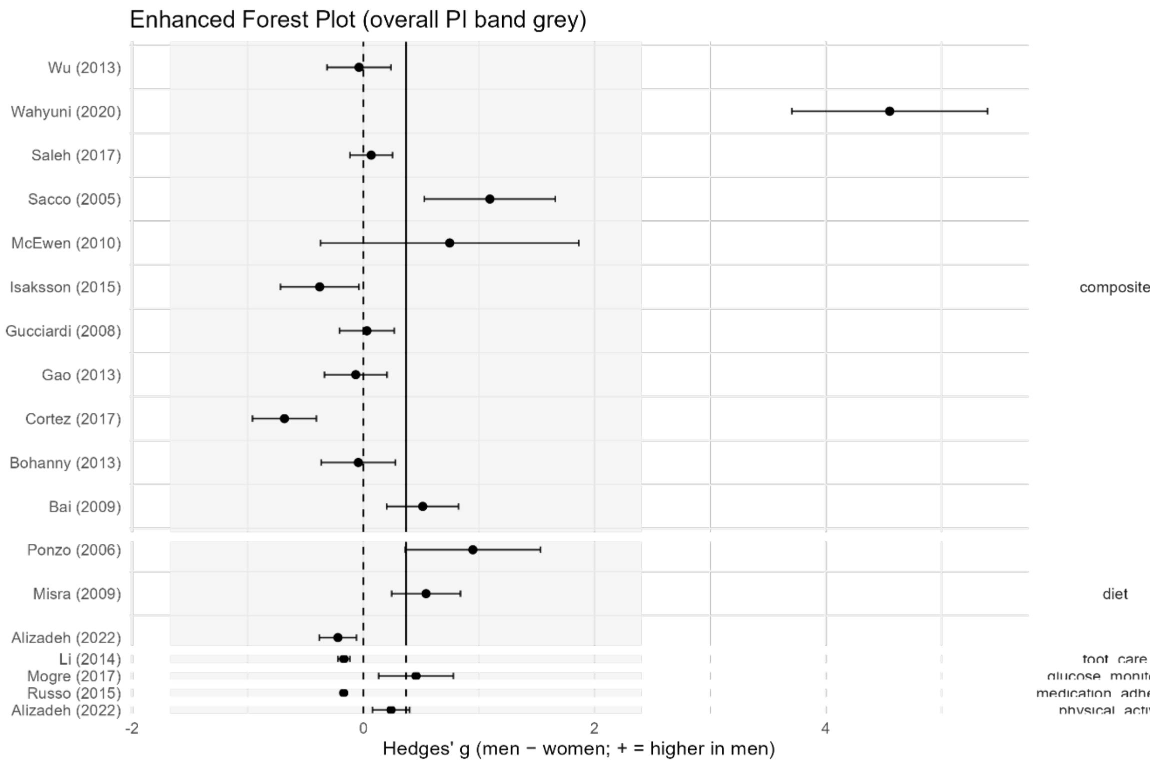


Figure 3. Enhanced forest plot by study and domain (overall prediction interval shaded); Study-level Hedges' g with 95% confidence intervals, grouped by behavioral domain; dashed vertical line shows the pooled estimate and the shaded band indicates the 95% prediction interval. Direction: positive values indicate higher scores in men.

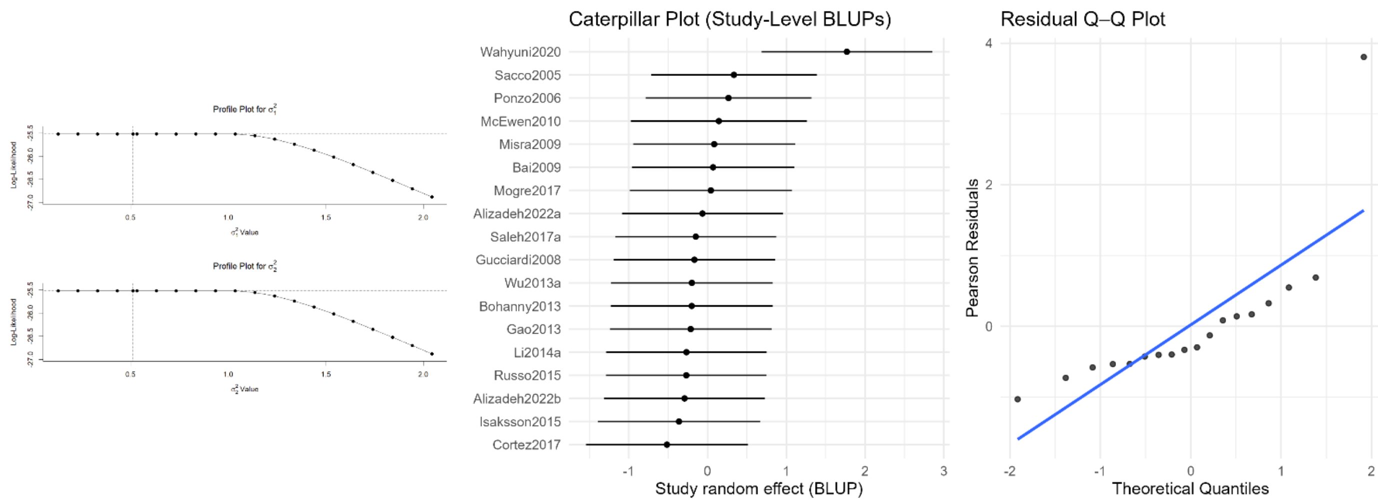


Figure 4. Heterogeneity dashboard: profile-likelihood curves, study BLUPs, and residual Q–Q; *Left:* profile-likelihood curves for the between- and within-study variance components (τ^2); curvature indicates identifiable nonzero heterogeneity. *Middle:* study-level BLUPs with intervals; points near zero indicate small random effects with a few positive deviations. *Right:* residual Q–Q plot; approximate linearity with an upper-tail deviation suggests a potentially influential positive residual to be examined in leave-one-out diagnostics.

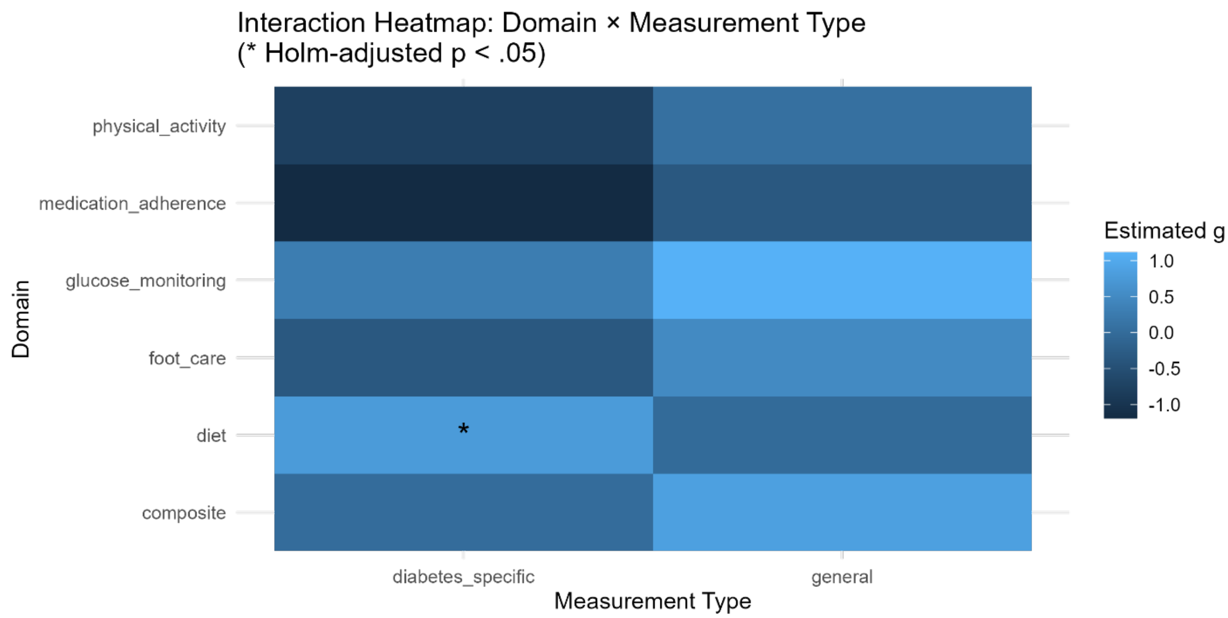


Figure 5. Interaction heatmap: domain × measurement type; Estimated domain × measurement-type contrasts on the Hedges' g scale; darker cells represent larger predicted differences. An asterisk denotes cells where simultaneous confidence intervals exclude zero. Sparse cells should be interpreted cautiously.

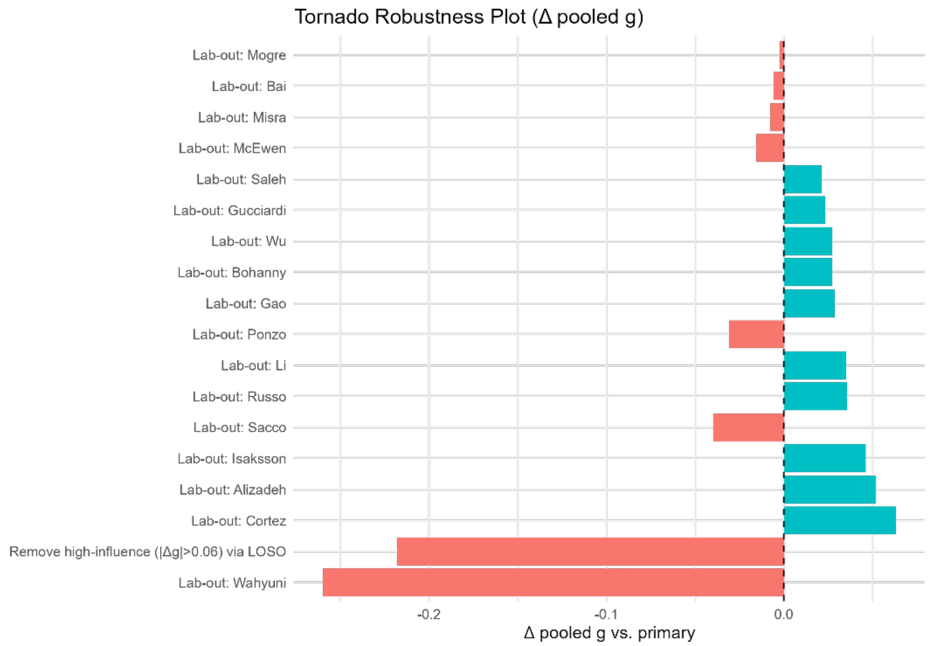


Figure 6. Tornado robustness plot (Δ pooled g for each sensitivity cut); Horizontal bars display the change in pooled Hedges' g relative to the primary model for each leave-one-lab-out scenario and for the aggregated removal of high-influence studies. Bars with the largest absolute Δ correspond to Wahyuni and Cortez; most other removals are near zero.

Online supplementary materials

Supplementary Material 1. Characteristics of included studies (random sample of 10).

Supplementary Material 2. Primary multilevel model (REML).

Supplementary Material 3. Moderator models (CR2 SEs): coefficients, p-values, ΔAICc, pseudo-R^2, and k.

Supplementary Material 4. Marginal effects by behavioral domain; Predicted Hedges' g with 95% simultaneous confidence intervals; ribbons reflect study weights. Intervals widely overlapping zero indicate limited evidence of domain differences at this stage.

Supplementary Material 5. Predicted Hedges' g versus year, faceted by domain; Bubble size is proportional to precision (1/SE). Slopes are near flat with wide bands in sparsely populated domains, consistent with the near-zero-year effect in Supplementary Material 4.

Supplementary Material 6. Contour-enhanced funnel plot with meta-analytic mean (blue); Scatter of Hedges' g versus precision (1/SE). The vertical blue line marks the pooled mean from the primary model. Asymmetry is visible with a relative excess of low-precision positive effects. Error note: one extreme low-precision positive point is flagged as a potential influence; its presence aligns with the significant Egger result reported in Table 4.

Supplementary Material 7. Small-study effects: Egger regression (precision on effect, Knapp–Hartung).

Supplementary Material 8. Sensitivity matrix (pooled Hedges' g and Δ vs. primary estimate).

Supplementary Material 9. Domain-wise pooled effects with 95% prediction intervals; Pooled Hedges' g by domain with 95% prediction intervals. Composite and diet show positive point estimates with wide intervals crossing zero. Empty rows denote domains without sufficient effects; no estimates were generated for those domains.

Supplementary Material 10. Summary of Findings (GRADE).