Personality Disorders: Theory, Research, and Treatment

Laboratory-Assessed Mistrust in Borderline Personality Disorder: A Meta-Analytic Review

Yağızcan Kurt, Megan Walker, Patrick Luyten, and Peter Fonagy

Online First Publication, May 22, 2025. https://dx.doi.org/10.1037/per0000739

REVIEW

Laboratory-Assessed Mistrust in Borderline Personality Disorder: A Meta-Analytic Review

Yağızcan Kurt[1, 2, 3], Megan Walker[4], Patrick Luyten[1, 2, 5], and Peter Fonagy[1, 2]
[1] Department of Clinical, Educational and Health Psychology, University College London
[2] Anna Freud, London, United Kingdom
[3] Department of Psychology, Istanbul Medeniyet University
[4] Division of Psychiatry, University College London
[5] Faculty of Psychology and Educational Sciences, University of Leuven

Although a recent systematic review examined the relationship between mistrust and borderline personality disorder (BPD; Preti et al., 2023), it did not statistically quantify the strength of this association. This highlights the need for a meta-analytic review of mistrust in BPD. To address this gap, we conducted a meta-analysis comparing laboratory-based assessments of mistrust between individuals with BPD and control groups. This meta-analysis was preregistered with International Prospective Register of Systematic Reviews (CRD42023479031). The final database search was conducted on October 17, 2024, across Embase, MEDLINE, PsycINFO, Scopus, and Web of Science to identify relevant studies. We included studies published in English or translated into English that employed valid, reliable laboratory-based measures of mistrust and psychometrically sound tools for assessing BPD symptoms or features. A three-level meta-analytic model was used to assess mistrust differences between BPD and control groups. The risk of bias in the included records was evaluated using the Joanna Briggs Institute Critical Appraisal Checklists. Our search yielded 1,717 reports, of which 26 met the inclusion criteria ($N = 3,716$). Based on 70 effect sizes from these reports, the meta-analysis indicated a small-to-moderate effect size ($g = 0.44$, 95% confidence interval = [0.27, 0.61], $p < .001$), demonstrating that individuals with BPD exhibited significantly higher levels of mistrust compared to controls. Subgroup analyses identified the mistrust paradigm as a statistically significant moderator. These results underscore the significant role of mistrust in BPD. We suggest that assessing mistrust and fostering interpersonal trust during treatment could lead to more effective interventions for BPD.

Keywords: mistrust, distrust, trust game, trustworthiness appraisal, borderline personality disorder

Supplemental materials: https://doi.org/10.1037/per0000739.supp

Borderline personality disorder (BPD) is a multifaceted mental health condition characterized by an unstable self-image, turbulent interpersonal relationships, instability in affect, and impulsive behaviors (American Psychiatric Association [APA], 2013). Prevalence estimates suggest that approximately 5.9% of the population is diagnosed with BPD (Grant et al., 2008), accounting for about 15%–20% of patients in both outpatient and inpatient settings (Gunderson et al., 2011). This significant prevalence underscores the profound impact of BPD

on individual well-being and its broader implications for public health (Oldham, 2006; Skodol et al., 2005).

The ability to build and maintain positive relationships is central to well-being. Accordingly, difficulties in interpersonal interactions can negatively impact mental health. The influence of interpersonal relationships on mental health highlights the importance of relational stability in the conceptualization and treatment of various psychopathologies (Howard et al., 2022). A notable example of a condition where interpersonal stability is particularly problematic is BPD (Gunderson, 2007).

Individuals with BPD often experience severe interpersonal difficulties, including extreme shifts between affection and contempt (Beck et al., 2015; Kernberg, 1967) and intense fears of real or perceived abandonment (APA, 2013). Trust is essential for fostering interpersonal vulnerability, building stable relationships, and establishing genuine connections based on the expectation of positive intentions from others (Lazarus et al., 2014). In individuals with BPD, the propensity for interpersonal difficulties is partly attributed to a deep-rooted mistrust of others, contributing to unstable relationships and conflict in social situations (Masland et al., 2020). While mistrust is a universal human experience, it tends to be significantly more intense in people with BPD (Preti et al., 2023). The *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (*DSM-5*) recognizes transient, stress-related paranoia, which implicitly reflects the mistrust associated with BPD (APA, 2013). Additionally, the Alternative Model for Personality Disorders criteria in the *DSM-5* (APA, 2013) suggest that unstable close relationships in BPD are frequently marked by mistrust.

Various theoretical frameworks explain the link between mistrust and BPD, consistently highlighting the perception of others as untrustworthy. Despite differing terminology, they converge on the role of temperament and early experiences in shaping these perceptions. Object relations theory emphasizes paranoid object relations (Kernberg, 1967), attachment theory highlights negative internal working models (Agrawal et al., 2004), schema theory addresses mistrust/abuse schemas (Young et al., 2003), and cognitive theory examines mistrust beliefs (Beck et al., 2015). These theories suggest that in hostile early environments, mistrust may function as a protective mechanism, shaping interpersonal relationships into adulthood. Additionally, mentalizing theory posits that difficulties in understanding mental states contribute to mistrust (Bateman & Fonagy, 2012). Deficits in mentalizing reduce epistemic trust—trust in communicated knowledge—hindering the assimilation of new information because of distorted perceptions of mental states (Fonagy et al., 2015). In BPD, this epistemic mistrust impairs the ability to recognize information as reliable, leading to rigid thought and behavior patterns (Fonagy et al., 2015).

Laboratory-based assessments of mistrust generally fall into two categories: paradigms that measure trust appraisals and those that assess trust-related behavior (Masland et al., 2020). Appraisal-based tasks typically involve participants evaluating the trustworthiness of others based on visual stimuli, such as photographs or short video clips depicting faces or social interactions. Participants are asked to rate how trustworthy they perceive the individuals to be. These paradigms are designed to capture trust biases—specifically, individuals' tendencies to perceive untrustworthiness in others, which are often shaped by longstanding beliefs and expectations about interpersonal relationships. In contrast, behavioral paradigms examine how participants act in situations that require trust. The most widely used example

is the trust game, along with its various adaptations. This task involves a multiround economic exchange between two players: an investor and a trustee. Each begins with a set amount of money. In a given round, the investor decides how much money to send to the trustee. The sent amount is then multiplied—typically tripled, though this can vary across studies—and transferred to the trustee. The trustee then chooses how much of this augmented amount to return to the investor. The game models reciprocal cooperation: larger investments signal greater trust, and generous returns reflect reciprocation. In addition to these exchange-based paradigms, other behavioral tasks assess how individuals integrate advice or feedback from others as a proxy for trust. Examples include the social valuation task and the judge–advisor system, which evaluate the extent to which individuals rely on or discount others' input when making decisions.

The empirical literature suggests an association between mistrust and BPD. Regarding self-reported mistrust, a meta-analysis found a small negative correlation between trust (as a facet of agreeableness in the Five-Factor Model) and BPD symptoms ($r = -.29$; Samuel & Widiger, 2008). Another reported a moderate association between BPD and the suspiciousness trait in the Alternative Model of Personality Disorders ($r = .46$; Watters et al., 2019). A third showed that individuals with BPD exhibited significantly higher mistrust/abuse schema scores than healthy controls (HCs; $g = 1.92$; Thimm & Chang, 2022).

Regarding laboratory-based assessed mistrust, research indicates that individuals with BPD are less likely to transfer monetary units in trust games (Niedtfeld & Kroneisen, 2020; Unoka et al., 2009) and tend to rate facial expressions as more untrustworthy (Miano et al., 2023). A recent systematic review also confirms that trust issues, primarily measured through laboratory-based tasks, are prevalent in individuals with BPD (Preti et al., 2023). However, while some studies report no significant association between mistrust and BPD (Scheunemann et al., 2023), others reveal paradoxical trust behaviors (Abramov et al., 2020). Given the significant role of mistrust in BPD symptoms and the variability in research findings, a meta-analytic approach assessing mistrust through laboratory-based tasks is essential to clarify this complex relationship.

Objectives

Contradictory findings in studies using laboratory-based tasks—some showing no relationship between mistrust and BPD, and others indicating that individuals with BPD are more trusting. These inconsistencies, particularly within laboratory-based paradigms, underscore the need for a meta-analysis to clarify the dynamics of mistrust measured through laboratory-based tasks.

Mistrust plays a significant role in the interpersonal difficulties observed in BPD (Lazarus et al., 2014). It also triggers core symptoms such as identity disturbance, mood instability, and impulsivity (Sharp, 2016) and is often linked to self-harm and suicidal behaviors following interpersonal conflict (Brodsky et al., 2006; Brown et al., 2002). Moreover, mistrust can undermine therapeutic alliances, leading to poorer treatment outcomes and higher dropout rates (Berghuis et al., 2021). Accordingly, a meta-analysis of mistrust differences is of clear clinical importance.

A systematic review on mistrust and BPD (Preti et al., 2023) exists but lacks statistical synthesis, highlighting the need for a meta-analytic approach to quantify the association between mistrust and BPD. To address this gap, we aim to review studies

using laboratory-based paradigms, such as trustworthiness appraisal tasks and trust games, within a meta-analytic framework.

This meta-analysis focuses specifically on laboratory-based assessments of mistrust. While self-report studies typically examine enduring or trait-like aspects of mistrust, laboratory paradigms are designed to capture how mistrust emerges in specific, situational contexts. By synthesizing findings from these paradigms, this review aims to highlight the contextual and potentially malleable nature of mistrust in BPD. Furthermore, self-report measures often assess explicit perceptions, whereas laboratory-based assessments tend to involve implicit measures—underscoring the importance of investigating mistrust through implicit approaches. Compared to self-report methods, laboratory tasks may better capture the influence of immediate contextual factors. A meta-analysis of such tasks may illuminate how shifting contexts shape mistrust and inform future research and clinical practice aimed at reducing mistrust in BPD.

Potential Moderators

Age (Clark & Eisenstein, 2013; Frías et al., 2017), sex (Bozzatello et al., 2024; van den Akker et al., 2020), and education (Frederiksen et al., 2016; Juurlink et al., 2022)—which have been associated with both mistrust and BPD in previous studies—were considered as potential moderators. In addition, we will examine study quality and publication year as potential moderators. Higher-quality studies and methodological advances over time may be associated with smaller effect sizes.

The type of task used to assess mistrust may also act as a moderator. Facial trustworthiness appraisal tasks rely on explicit ratings, whereas trust games involve implicit measures. Methodological differences between these tasks may lead to varying results. Additionally, the diversity of instruments used to assess BPD could contribute to heterogeneity across studies.

Recent research suggests that the paradigm's intended effect—whether inducing mistrust, trust, or neutrality—may act as a moderator. Paradoxical trust behaviors in BPD have been observed, with trust decreasing in socially accepting contexts (Liebke et al., 2018) and increasing after deception but declining with cooperative partners (Abramov et al., 2020). We aim to examine how these conditions affect mistrust levels in BPD versus controls. This will clarify whether trust-enhancing cues (e.g., eye contact and smiling) reduce mistrust or if negative interactions paradoxically increase trust, signaling heightened vulnerability to exploitation.

Method

Eligibility Criteria

The inclusion criteria for studies were as follows: (a) provided original quantitative data, (b) published in a peer-reviewed journal, (c) used a validated laboratory-based task to measure mistrust, (d) employed a psychometrically validated and reliable instrument to diagnose BPD or assess BPD traits, (e) compared levels of mistrust between individuals diagnosed with BPD or exhibiting high BPD traits and a comparison group (e.g., HCs, individuals with low BPD traits, or clinical controls [CCs]) or examined the cross-sectional association between mistrust and BPD, and (f) included sufficient data to calculate effect sizes (e.g., sample size, mean, and standard deviation).

The following studies were excluded: (a) books, book chapters, theses, and conference abstracts; (b) commentaries, editorials, and letters to the editor; (c) qualitative research; (d) single case studies and case series; (e) review articles and other meta-analyses; (f) unpublished articles or articles not translated into English; and (g) unpublished articles. No restrictions were placed on the year of publication.

Information Sources

The final database search was completed by Yağızcan Kurt on October 17, 2024, across Embase (via Ovid, 1974–present), MEDLINE (via Ovid, 1946–present), PsycINFO (via Ovid, 1806–present), Scopus (1788–present), and Web of Science (1900–present).

A decision was made after the initial analysis to examine whether the inclusion of gray literature influenced the meta-analytic results. Accordingly, we searched APA PsycExtra (via Ovid, 1908–present) and ProQuest Dissertations & Theses Global.

Search Strategy

The search strategy was developed by Yağızcan Kurt in consultation with a subject liaison specialist in biological sciences and psychology, who is a librarian. Its accuracy was cross-checked by assessing the strategy's ability to identify studies from a recent systematic review (Preti et al., 2023). For each concept, where applicable, keywords were combined with database-specific subject headings. We used the keywords trust*, mistrust*, untrust*, distrust, and cooperation to capture mistrust-related concepts and BPD for BPD-related studies. Detailed search strategies for all databases are provided in Supplemental Table 1 in the online supplemental materials.

Selection Process

All database records were imported into EndNote 20 for duplicate removal. Yağızcan Kurt screened titles and abstracts using predefined criteria, while Megan Walker independently screened a random 25% subsample. Interrater reliability was assessed, and discrepancies were resolved collaboratively. Yağızcan Kurt then reviewed all eligible full texts; Megan Walker reviewed another random 25% subsample. Interrater reliability was reassessed, and disagreements were resolved through discussion.

Data Collection Process

A data extraction form was created in Microsoft Excel. Yağızcan Kurt extracted data from all included records; Megan Walker independently extracted a random 25% subsample. Extracted data were compared, interrater reliability was assessed, and discrepancies were resolved through discussion.

Data Items

Effect sizes were extracted from studies with multiple samples. When both subsample and total sample effect sizes were reported, subsample values were prioritized. For studies reporting multiple measures of the same construct (e.g., different BPD assessments), an effect size was extracted for each. In cases of duplicate samples, only one effect size was used. Extracted data included: (a) study details (title, authors, year, journal); (b) sample characteristics for the full sample, BPD group, and controls (sample size, mean age, % female, mean education); (c) setting (e.g., students, inpatients, and outpatients); (d) research design (case–control or correlational);

(e) mistrust paradigm (e.g., trust game and facial trustworthiness task); (f) BPD measure (e.g., Structured Clinical Interview for *DSM-IV* Axis II Personality Disorders and Borderline Symptom List-23); (g) quality scores; and (h) effect size.

Study Risk of Bias Assessment

Risk of bias was assessed using the Joanna Briggs Institute Critical Appraisal Checklists for Case Control Studies and Analytical Cross-Sectional Studies (Moola et al., 2020). The Case Control Checklist includes 10 items: (a) comparability of groups, (b) matching, (c) objective criteria for identifying cases and controls, (d) validity and reliability of exposure measures, (e) consistency in measuring exposure for cases and controls, (f) identification of confounding factors, (g) strategies to address confounding, (h) validity and reliability of outcome measures; (i) duration of exposure, and (j) appropriateness of statistical analysis. The item on exposure duration was excluded as it was not relevant to our meta-analytic objectives.

The Cross-Sectional Checklist includes eight items: (a) clarity of inclusion criteria; (b) description of study subjects and setting; (c) validity and reliability of exposure measures; (d) use of objective, standardized criteria to measure conditions; (e) identification of confounding factors; (f) strategies to address confounding; (g) validity and reliability of outcome measures; and (h) appropriateness of statistical analysis.

Quality assessment was conducted at the level of extracted effect sizes rather than studies because of variations in BPD measures and the presence of both case–control and correlational designs within some studies. Each effect size was rated as "yes," "no," "uncertain," or "not applicable." Quality scores were calculated as the percentage of "yes" responses out of applicable items. Yağızcan Kurt assessed all effect sizes; Megan Walker independently reviewed a random 25% subsample. Interrater reliability was assessed, and discrepancies were resolved through discussion.

Effect Measure

Given the expected variability in mistrust between BPD and comparison groups, we used standardized mean differences as our effect size measure, specifically Hedges' *g* to correct for small sample bias (Hedges, 1981). Higher *g* values indicate greater group differences in mistrust. We interpreted effect sizes using Cohen's (1988) benchmarks: .20 (small), .50 (medium), and .80 (large). When *g* was not reported, we calculated it from available data (sample sizes, means, and standard deviations). For measures framed in terms of trust, we reversed scores as needed to align with the direction of mistrust. Finally, to convert Pearson correlation coefficients to Hedges' *g*, we used a pragmatic approximation. Since the standard *r-to-d* formula was developed for point-biserial correlations, we applied it assuming a hypothetical group contrast along the continuous variable. This may inflate effect sizes when both variables are continuous (McGrath & Meyer, 2006). When studies reported only the total sample size, we assumed equal group sizes; otherwise, we used the reported group sizes. To assess the impact of this decision, we conduct a sensitivity analysis excluding converted correlational effect sizes.

Synthesis Methods

We employed a three-level meta-analytic model (Assink & Wibbelink, 2016; Harrer et al., 2021) instead of the conventional two-level model to account for dependencies among effect sizes derived from the same studies, which are likely to be more similar than those from different studies. This model considers three sources of variance: (a) sampling variance of effect sizes (Level 1), (b) within-study variance (Level 2), and (c) between-study variance (Level 3; Cheung, 2014; Van den Noortgate et al., 2015).

Analyses were conducted using R Version 4.3.1 (R Core Team, 2023) with a three-level random-effects model, employing restricted maximum-likelihood estimation with Knapp-Hartung adjustment (Harrer et al., 2021; Knapp & Hartung, 2003). We utilized the rma.mv function from the metafor package Version 4.2.0 (Viechtbauer, 2010).

To assess heterogeneity, we conducted one-sided log-likelihood ratio tests to evaluate the significance of within-study (Level 2) and between-study (Level 3) variances. According to Hunter and Schmidt's (1990) criteria, heterogeneity was considered significant if less than 75% of the variance originated from sampling variance (Level 1). Additionally, we applied Cheung's (2014) formula to evaluate the distribution of variance across the three levels of the meta-analytic model.

After identifying significant heterogeneity, we conducted moderation analyses to examine the influence of specific factors on the relationship between mistrust and BPD. For categorical moderators, we generated dummy variables using the fastDummies package Version 1.7.4 (Kaplan, 2024). Continuous moderators were centered around their means before testing their effects on effect size. The significance of potential moderators was evaluated through an omnibus test, with each moderator differing in the number of levels. For example, BPD assessment methods included (a) interview and (b) self-report; the direction of the trust paradigm included (a) trust enhancing, (b) mistrust enhancing, and (c) neutral and the type of trust paradigm comprised (a) trustworthiness appraisal and (b) trust behavior.

We performed a sensitivity analysis to determine whether incorporating gray literature affects the study's overall conclusion. Although this analysis was not initially outlined in the preregistration, it was added later to adhere to best practices for meta-analyses. In addition, we conducted a sensitivity analysis to assess whether excluding effect sizes converted from correlations to standardized mean differences affected the study's overall conclusions. As our approach prioritized data inclusion over strict statistical equivalence, we sought to test the impact of this decision to ensure the robustness of our findings.

Reporting Bias Assessment

To assess reporting bias, we used multiple approaches. Funnel plots adapted for the three-level meta-analytic model displayed effect sizes against standard errors to detect selective outcome reporting bias (Fernández-Castilla et al., 2020). Additionally, a separate funnel plot illustrated study effects, calculated via individual random-effects meta-analyses, accounting for variability, outcomes reported, and sample size.

Since traditional publication bias methods do not handle dependent effect sizes well (Rodgers & Pustejovsky, 2021), we applied two regression techniques: the Egger multilevel meta-analysis (MLMA) test and the Egger–Sandwich Test. These analyses, conducted using metafor (Viechtbauer, 2010) and robumeta

(Fisher et al., 2023), integrate Egger's regression with MLMA and robust variance estimation, effectively addressing dependencies and selective reporting.

Transparency and Openness

We adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 guidelines for systematic reviews (Page et al., 2021). All data, analysis code, and research materials are available and can be found as the additional online materials (https://osf.io/t6syu/?view_only=c07a56eb926 64d31bd4cf235e1b3ffe3). This review was preregistered with the International Prospective Register of Systematic Reviews (https://www.crd.york.ac.uk/PROSPERO/view/CRD42023479031).

Although not preregistered, we conducted two sensitivity analyses in line with best practices for meta-analyses to assess how the inclusion of gray literature and the exclusion of converted correlational effect sizes impact the robustness of our results. Additionally, although we intended to analyze the effect size of the difference between BPD and CCs, the limited number of available effect sizes prevented us from conducting a separate meta-analysis or moderator analysis.

Results

Study Selection (Flow of Studies)

Our search initially identified 1,717 records. After removing 784 duplicates via EndNote, 933 records remained for title and abstract screening. We excluded 876 at this stage, leaving 57 for full-text review. Of these, 27 were excluded for not meeting inclusion criteria. From the remaining 30, we included 25 directly and contacted authors for the remaining five. One author provided data, resulting in 26 included records. A full list of included records is available in Supplemental Table 2 in the online supplemental materials.

Throughout the selection process, Megan Walker independently reviewed a randomly selected 25% subsample at each stage. The initial interrater agreement for abstract screening was moderate ($\kappa = 0.53$, $SE = 0.12$, 95% confidence interval [CI] = [0.29, 0.78]). To enhance agreement, we revised the inclusion and exclusion criteria and reassessed a randomly selected subsample. This led to an improvement in interrater reliability ($\kappa = 0.74$, $SE = 0.10$, 95% CI = [0.53, 0.94]), though it remained below the desired threshold of .80. We conducted another round of revisions, further increasing agreement ($\kappa = 0.81$, $SE = 0.08$, 95% CI = [0.66, 0.96]). For full-text screening, interrater agreement was also assessed but was suboptimal ($\kappa = 0.70$, $SE = 0.19$, 95% CI = [0.32, 1.00]). To improve agreement, we refined the criteria again and reassessed another random subsample, achieving an interrater agreement of $\kappa = 0.87$, $SE = 0.13$, 95% CI = [0.62, 1.00]. All discrepancies were discussed and resolved. Figure 1 provides a flowchart illustrating the study selection process.

Study Selection (Excluded Studies)

During the full-text screening stage, we excluded 31 studies from the meta-analysis. The reasons for these exclusions are detailed in Figure 1, and a comprehensive reference list for each excluded study is provided in Supplemental Table 3 in the online supplemental materials.

Study Characteristics

The characteristics of all studies included in the meta-analysis are provided in Supplemental Table 4 in the online supplemental materials.

Risk of Bias in Studies

The initial interrater reliability for the risk of bias assessment was suboptimal ($\kappa = 0.69$, $SE = 0.06$, 95% CI = [0.57, 0.82]). Following a review of the discrepant items, we selected a new 25% subsample of the records for reassessment. This led to a marked improvement in interrater reliability ($\kappa = 0.84$, $SE = 0.09$, 95% CI = [0.67, 1.00]). The quality of the records varied: effect sizes for cross-sectional association studies ranged from 4/8 to 6/8, while those for case–control studies ranged from 4/9 to 9/9. The majority of cross-sectional association effect sizes scored 5 out of 8 (41.7%) and 6 out of 8 (41.7%), and most case–control effect sizes scored 5 out of 9 (27.4%).

All correlational studies used valid and reliable measures to assess BPD and mistrust, along with appropriate statistical analyses. However, 91.7% relied on self-reported assessments rather than other diagnostic tools such as structured interviews. Moreover, 58.3% were based on samples with unclear inclusion/exclusion criteria, and 66.7% did not report how confounding variables were handled. Supplemental Table 5 in the online supplemental materials details the quality assessment for each cross-sectional study. As results were consistent within study types, the information is consolidated by study rather than effect size.

All included case–control studies used valid, reliable measures for mistrust and BPD and employed appropriate statistical analyses, meeting our inclusion criteria. Most demonstrated group comparability (68%) and applied consistent criteria for identifying cases and controls (83.9%). However, 74.2% of the 62 case–control effect sizes did not adequately match cases and controls, and 51.6% did not report how confounding variables were addressed. As with the correlational studies, Supplemental Table 6 in the online supplemental materials presents results by study rather than effect size.

Only one study, Fineberg et al. (2018), achieved a perfect quality score, with all four of its effect sizes scoring 9 out of 9. A summary of the quality assessment is shown in Supplemental Figures 1–4 in the online supplemental materials.

Results of Individual Studies

The analysis yielded 57 positive and 13 negative effect sizes. A forest plot was created to illustrate the weight and precision of each study within the meta-analysis (see Figure 2). For a comprehensive overview of all included studies and their characteristics, refer to Supplemental Table 4 in the online supplemental materials. The detailed descriptions of the laboratory tasks used to measure mistrust for each effect size are provided in Supplemental Table 7 in the online supplemental materials.

The largest effect sizes ($g > 0.80$) were primarily linked to trustworthiness appraisal tasks, with only two from trust games. In facial appraisal studies, individuals with BPD rated neutral faces as more untrustworthy than controls (Biermann et al., 2022; Fertuck et al., 2019, 2023; Nicol et al., 2013). In trust game predictions, individuals with BPD expected worse outcomes (Unoka et al., 2009).

Figure 1

PRISMA Flowchart Illustrating the Study Selection Process
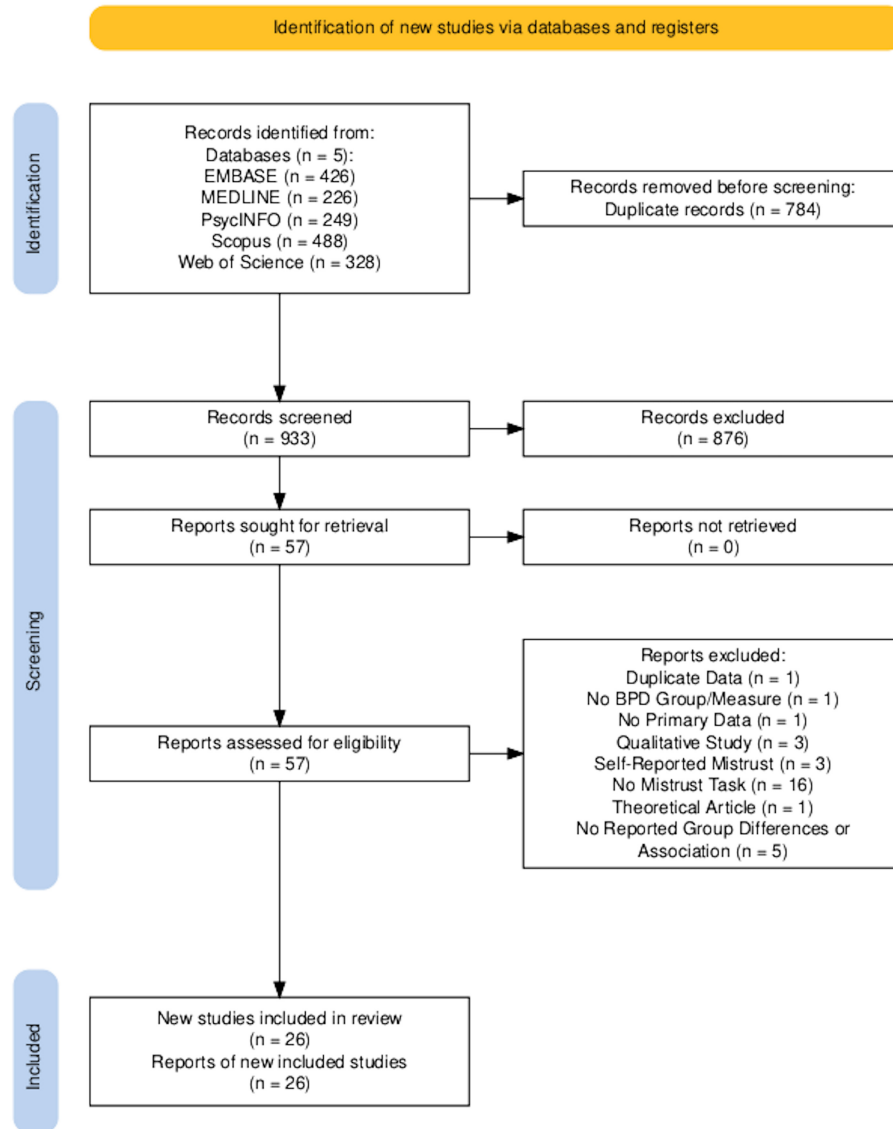
Note. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses. See the online article for the color version of this figure.

In irony detection tasks, they rated praising remarks as less sincere (Felsenheimer et al., 2022). Additional trust game findings showed reduced trust in BPD, especially after intranasal oxytocin and with unattractive partners (Ebert et al., 2013).

Thirteen effect sizes were negative, challenging the assumption of heightened mistrust in BPD. Except for one, all were from trust behavior measures. In trust games, individuals with BPD were more trusting than controls when their partner was attractive (Ebert et al., 2013), after acetaminophen administration (Roberts et al., 2018), or in adolescent samples (Sharp et al., 2023). Findings from judge–advisor tasks, social valuation tasks, and implicit irony detection suggest greater trust in BPD groups, as they were more likely to accept advice and interpret messages less negatively (Fineberg et al., 2018; Scheunemann et al.,

2023; Shapiro-Thompson et al., 2023). Notably, implicit assessments in the irony task showed higher trust among BPD participants, contrasting with explicit ratings (Felsenheimer et al., 2022).

Results of Synthesis (Characteristics of Contributing Studies)

Sample Characteristics

A total of 29 independent samples were identified, comprising a pooled sample of 3,716 participants ($n = 1,073$ for BPD groups, $n = 2,458$ for control groups). The pooled mean age was 30.14 years ($SD = 8.82$) for BPD participants and 29.39 years ($SD = 9.42$) for controls. In the BPD groups, 71.46% of participants were female compared to 69.04% in the control groups. The pooled

Figure 2

Forest Plot Displaying All Effect Sizes Included in the Meta-Analysis
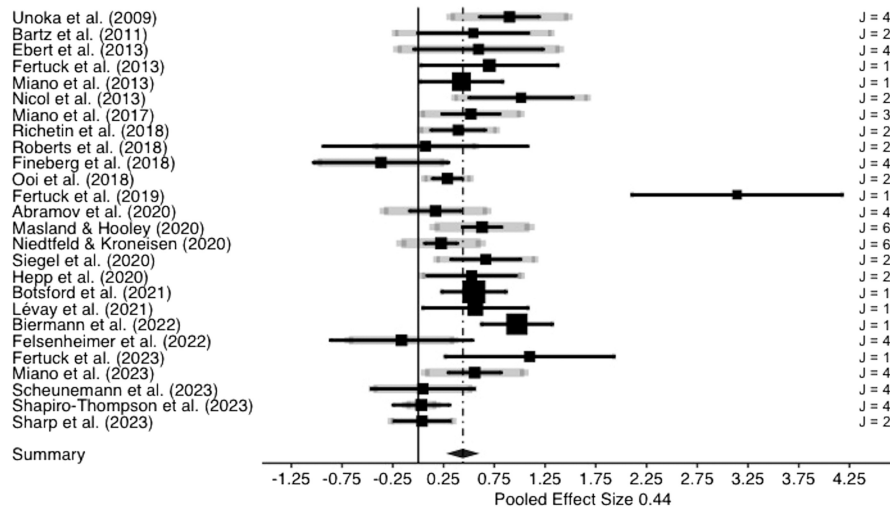
mean number of years of education was 13.62 ($SD = 1.65$) for BPD participants and 14.09 ($SD = 0.59$) for controls.

Publication Characteristics

The majority of the research was conducted in the United States ($n = 10$) and Germany ($n = 9$), followed by the United Kingdom ($n = 3$), Hungary ($n = 2$), Italy ($n = 1$), and Australia ($n = 1$). The most frequent publication years were 2020 ($n = 5$), 2023 ($n = 5$), 2013 ($n = 4$), and 2018 ($n = 5$), spanning from 2009 to 2023. The studies were published in 15 different journals, with the most common being the *Journal of Personality Disorders* ($n = 8$) and *Borderline Personality Disorder and Emotional Dysregulation* ($n = 3$).

Study Design and Measures

The majority of effect sizes were derived from studies comparing mistrust between BPD patients and HCs ($n = 51$). Only four compared BPD patients to CCs and seven compared individuals with high versus low BPD traits. Out of the 70 effect sizes obtained, 62 employed a case–control design, and eight utilized a correlational design. The most frequently used measures for assessing mistrust were facial trustworthiness tasks ($n = 19$), trust games ($n = 15$), and social valuation tasks ($n = 8$). The Structured Clinical Interview for *DSM-IV* Axis II Personality Disorders interview was the most widely used method for measuring BPD symptoms and severity, featuring in 31 effect sizes; other methods included the International Personality Disorder Examination ($n = 9$), the schedule for nonadaptive and adaptive personality-2nd edition-BPD ($n = 6$), the McLean screening instrument for BPD ($n = 4$), and the revised diagnostic interview for borderlines ($n = 4$). Of the 70 effect sizes analyzed, 50 used interviews to assess BPD and 20 used self-report measures. Forty-one of these effect sizes assessed trustworthiness appraisals, while 29 focused on mistrust behavior. Regarding the intended effect of the design on mistrust, 37 of the effect sizes employed neutral conditions, 19 used trust-enhancing conditions, and 14 utilized mistrust-enhancing conditions.

Results of Synthesis (Statistical Synthesis Results)

Based on a three-level meta-analysis of 26 studies comprising 70 effect sizes, the pooled standardized mean difference in mistrust between individuals with BPD and control groups was found to be $g = 0.44$ (95% CI = [0.27, 0.61], $p < .001$). The estimated variance components were $\tau^2 = .06$ at Level 2 and $\tau^2 = .13$ at Level 3.

Results of Synthesis (Investigations of Heterogeneity)

An initial test for heterogeneity revealed significant differences among effect sizes, $Q(70) = 313.10$, $p < .001$. However, this result does not fully capture the nature of heterogeneity in a three-level meta-analytic model, as it does not distinguish between within-study and between-study variance.

To address this, we conducted two separate log-likelihood ratio tests to assess the significance of within-study and between-study variance (τ^2 at Level 2 = .06 and τ^2 at Level 3 = .13, respectively). For within-study variance, we compared the fit of the three-level model with that of a two-level model excluding within-study variance. The three-level model showed a significantly better fit ($\chi^2 = 16.48$, $p < .001$), indicating significant within-study heterogeneity. Similarly, to assess between-study variance, we compared the three-level model with a two-level model that excludes between-study variance. Again, the three-level model demonstrated a significantly better fit ($\chi^2 = 12.05$, $p < .001$), indicating significant heterogeneity between studies.

Following the approach of Cheung (2014), we examined the distribution of variance across levels. The analysis revealed that sampling error variance (I^2 at Level 1) accounted for 8.96% of the total variation, within-study variance (I^2 at Level 2) for 29.07%, and between-study variance (I^2 at Level 3) for 61.97%. Since only a small portion of the total variation was attributed to sampling error, there is considerable variance among the effect sizes. This underscores the need for a moderator analysis based on the criteria of Hunter and Schmidt (1990).

Results of Synthesis (Moderators)

The type of mistrust paradigm used significantly moderated the effect size, $F(1, 68) = 8.06$, $p = .006$. Both trustworthiness appraisals ($k = 41$, $g = 0.56$, 95% CI = [0.38, 0.74], $p < .001$) and mistrust behavior measures ($k = 29$, $g = 0.21$, 95% CI = [0.00, 0.42], $p = .046$) yielded statistically significant effect sizes, with trustworthiness appraisals showing a larger effect size than mistrust behavior measures.

The direction of the mistrust paradigm—whether neutral, mistrust enhancing, or trust enhancing—did not significantly influence the overall effect size, $F(2, 67) = 1.64$, $p = .202$. Significant differences in mistrust between BPD and control groups were observed under neutral conditions ($k = 38$, $g = 0.48$, 95% CI = [0.30, 0.66], $p < .001$), mistrust-enhancing conditions ($k = 14$, $g = 0.47$, 95% CI = [0.22, 0.73], $p < .001$), and trust-enhancing conditions ($k = 19$, $g = 0.27$, 95% CI = [0.03, 0.52], $p = .030$); however, there were no significant differences in effect sizes among these conditions.

The method of BPD assessment did not significantly affect the overall effect size, $F(1, 68) = 0.14$, $p = .705$. Studies using diagnostic interviews yielded a statistically significant effect size ($k = 50$, $g = 0.46$, 95% CI = [0.26, 0.65], $p < .001$) as did studies employing self-report methods to assess BPD traits ($k = 20$, $g = 0.40$, 95% CI = [0.12, 0.68], $p = .006$). No significant difference was found between the effect sizes derived from interviews and self-reports.

All other potential moderators—including age ($k_{BPD} = 65$, $k_{Controls} = 63$), education level ($k_{BPD} = 14$, $k_{Controls} = 14$), gender ($k_{BPD} = 57$, $k_{Controls} = 55$), quality assessment scores ($k = 70$), and publication year ($k = 70$)—were found to have no significant association with the effect size.

Results of Synthesis (Sensitivity Analyses)

We conducted sensitivity analyses to assess the robustness of our findings. First, we included effect sizes from gray literature in a three-level model (28 studies, 73 effect sizes), which yielded a pooled effect of $g = 0.41$, 95% CI = [0.25, 0.57], $p < .001$; $\tau^2 = .05$ (Level 2), $\tau^2 = .12$ (Level 3). This closely aligned with the original analysis ($g = 0.44$, 95% CI = [0.27, 0.61], $p < .001$; $\tau^2 = .06$, $\tau^2 = .13$), suggesting minimal influence of gray literature.

A second analysis excluded effect sizes converted from Pearson's r to Hedges' g. The resulting model (24 studies, 62 effect sizes) yielded $g = 0.44$, 95% CI = [0.25, 0.63], $p < .001$, with $\tau^2 = .09$ (Level 2) and $\tau^2 = .13$ (Level 3). Estimates remained consistent, suggesting these conversions did not meaningfully influence the overall findings.

Reporting Biases

Funnel plots were generated using an extension of the three-level meta-analytic model to visualize all effect sizes (see Supplemental Figure 6 in the online supplemental materials) and the mean effects across studies (see Supplemental Figure 7 in the online supplemental materials). Although both plots included an outlier, the distribution at the study level appeared relatively symmetrical.

Since conclusions cannot be drawn solely from visual inspection, additional statistical analyses were conducted. The Egger–Sandwich test revealed a significant result ($\beta = 2.96$, $SE = 1.49$, $p = .041$), indicating potential publication bias. However, the Egger MLMA test did not yield a significant result ($\beta = 2.22$, $SE = 0.77$, $p = .11$), suggesting that publication bias may not be a significant concern.

Discussion

Summary of Findings

We conducted a three-level meta-analysis of 70 effect sizes from 26 studies examining mistrust in individuals with BPD compared to control groups. The analysis identified a small-to-medium but significant effect size (Hedges' $g = 0.44$), indicating that individuals with BPD exhibit higher levels of mistrust.

Significant variability was found both within and between studies. This justified the use of a three-level meta-analysis to account for both within-study and between-study variance, as well as moderator analyses to examine potential factors contributing to the observed heterogeneity. Moderator analyses revealed that factors such as age, gender, and education level did not significantly influence the overall effect size. Similarly, variables like publication year, study quality, the direction of the trust paradigm, and the method of BPD assessment were not significantly associated with the results.

However, the type of measurement used to assess mistrust significantly moderated the effect size. Specifically, trustworthiness appraisal tasks were associated with a more pronounced difference in mistrust levels compared to behavioral mistrust tasks.

Finally, the sensitivity analyses—including those that incorporated effect sizes from gray literature and those that excluded correlational effect sizes—did not alter the overall interpretation of the meta-analytic results.

Interpretation of Main Finding

The increased mistrust observed in individuals with BPD suggests that mistrust may play a role in the disorder, potentially shaping their appraisal styles and behaviors. This predisposition toward mistrust likely contributes significantly to the interpersonal challenges associated with BPD, hindering the formation and maintenance of healthy relationships.

Based on our results from trustworthiness appraisal tasks, individuals with BPD perceive neutral faces as untrustworthy, which could lead them to avoid trustworthy individuals and, in turn, miss out on positive, trust-based relationships. Conversely, mistrustful behavior—evidenced in trust games, for instance—may be perceived by others as a lack of cooperation, potentially leading to negative reactions and fostering interpersonal conflict. Consequently, mistrustful appraisals and behaviors can lead to interpersonal dysfunction and dissatisfaction in relationships. These strained interactions may worsen other severe symptoms of BPD, such as impulsivity and suicidal behaviors in response to interpersonal issues (Sharp, 2016). Thus, mistrust may contribute to a cycle involving core symptoms of BPD and could be associated with various symptom domains, including relationship difficulties, identity problems, affective instability, and impulsivity. However, as we examined BPD as a whole rather than specific symptom domains, any discussion regarding the role of mistrust across these domains remains speculative and warrants further investigation.

Our finding that individuals with BPD exhibit higher levels of mistrust compared to controls is consistent with recent research, including a systematic review (Preti et al., 2023) and previous

meta-analyses of self-reported mistrust in BPD (Samuel & Widiger, 2008; Thimm & Chang, 2022; Watters et al., 2019). Preti et al. (2023) identified five studies showing consistent trust appraisal deficits in individuals with BPD, with most studies in their analysis also reporting reduced trustful behavior, aligning with our findings. Similarly, meta-analyses of self-reported mistrust in BPD have reported effect sizes ranging from nearly moderate ($r = .29$) to large ($g = 1.92$; Samuel & Widiger, 2008; Thimm & Chang, 2022; Watters et al., 2019). However, our study identified a smaller effect size. One possible explanation is that, unlike the previous meta-analyses, our analysis exclusively included laboratory-based measures of mistrust. This approach may mitigate the overreporting or underreporting of mistrust and BPD symptoms often seen in self-report methods, resulting in a more conservative estimation of effect sizes.

Interestingly, the discrepancy between self-reported and laboratory-assessed mistrust suggests that individuals with BPD may not consistently exhibit behavioral signs of mistrust, despite endorsing it in self-reports. Two studies in our meta-analysis observed mismatches between reported trust beliefs and actual trust behaviors (Sharp et al., 2023; Unoka et al., 2009). This points to the possibility that real-time trust decisions are less influenced by conscious beliefs and more by immediate context, with mistrust fluctuating rather than reflecting stable dispositions.

Another possible explanation for the smaller effect size is that laboratory-based measures in psychology often lack the rigorous psychometric validation typically applied to self-report instruments. For example, self-reported and behavioral measures of impulsivity show only minimal overlap (Cyders & Coskunpinar, 2011). A similar pattern may explain the inconsistencies reported by Sharp et al. (2023) and Unoka et al. (2009), where self-reported mistrust did not align with laboratory-assessed mistrust. This suggests that the weaker effect sizes may reflect limitations in the measurement of behavioral mistrust rather than a genuinely weaker association.

Alternatively, self-reported mistrust may be more reflective of real-world experiences than laboratory-based assessments, serving as a more direct measure of mistrust. Laboratory tasks might not prompt participants to deeply consider their trust beliefs in relation to personality traits (Brenner & DeLamater, 2016), potentially overlooking a stronger link between mistrust and BPD because of their less explicit and more variable nature. For example, trust games might be perceived primarily as measures of cooperative behavior rather than trust itself.

In *DSM-5*, transient paranoid ideation is a symptom of BPD, often involving ideas of reference, persecution, and delusions (Freeman & Garety, 2000). This differs from generalized, nondelusional interpersonal mistrust, which in BPD is more pervasive than transient. The tasks in this meta-analysis assess mistrust rather than paranoia, as they do not measure delusions or severe mistrust. Instead, they capture a more stable, reality-aligned form of interpersonal mistrust. Evidence from facial trustworthiness tasks and self-report studies suggests that mistrust in BPD is a persistent trait rather than a fleeting state. The Five-Factor Model of Personality Disorders also includes "distrust" in the BPD trait profile (Mullins-Sweatt et al., 2012), reinforcing this distinction. Recognizing this could refine diagnostic criteria by accounting for persistent, yet nondelusional, mistrust in BPD.

The significant mistrust observed in individuals with BPD aligns with current theoretical frameworks on mistrust in BPD, such as attachment theory (Bowlby, 1980), object relations theory (Kernberg et al., 2008), cognitive theory (Beck et al., 2015), and

schema theory (Young et al., 2003). These models emphasize the impact of adverse early-life experiences on the development of mistrustful mental representations, with numerous studies confirming strong associations between BPD symptom severity and childhood trauma (Yuan et al., 2023). Nonetheless, our findings also suggest that later environmental influences may affect these mistrust representations.

Furthermore, our findings support mentalizing-based explanations of mistrust (Fonagy et al., 2015). Specifically, individuals with BPD perceive neutral faces as more untrustworthy than controls, reflecting hypermentalizing tendencies (Sharp & Vanwoerden, 2015), where mental state inferences extend beyond facial expressions alone. This suggests that overinterpreting social cues may contribute to heightened mistrust toward others.

Studies that identified the largest disparities in mistrust between BPD and control groups primarily employed trustworthiness appraisal tasks, supporting the notion that mistrust beliefs, rather than behaviors, are prominent in BPD. Furthermore, in one study showing large disparities between groups, the tendency to perceive an unattractive partner as untrustworthy may indicate a dysfunctional assessment of mistrust in individuals with BPD, as attractiveness may not reliably predict trustworthiness (Ebert et al., 2013).

Upon analyzing the negative effect sizes, we observed that in tasks such as the social valuation task, judge–advisor task, and irony detection task, individuals with BPD demonstrated greater trust in communicated or social information compared to controls. This challenges the epistemic mistrust hypothesis in BPD and underscores the need for further investigation, as much of the discussion has remained theoretical. One possible explanation is that individuals with BPD may not have difficulty trusting information that lacks emotional salience or personal relevance. In other words, they may struggle more with affective-level learning than with cognitive-level learning, which requires further testing through experimental designs.

Additionally, negative effect sizes from some trust games support the view of potential dysfunction in assessing cues related to trustworthiness. For example, individuals with BPD were more trusting than controls when their game partner was perceived as attractive (Ebert et al., 2013). Furthermore, the administration of acetaminophen led to increased trust in individuals with BPD, a phenomenon the authors explained as extreme shifts from an all-bad to an all-good representation (Roberts et al., 2018). This helps explain the unexpected finding that individuals with BPD were more trusting than HCs.

Interpretation of Moderators

Our analysis revealed that age did not significantly moderate the relationship between mistrust and BPD. While prior research indicates that both mistrust and the severity of BPD symptoms tend to decrease with age (Clark & Eisenstein, 2013; Frías et al., 2017), this effect might be balanced by an increase in negative life events or the stabilization of one's environment over time. These opposing factors could neutralize each other, explaining why age did not emerge as a significant moderator in our study.

Similarly, gender was not a significant moderator despite evidence from previous studies suggesting a higher prevalence of mistrust and BPD in women compared to men (APA, 2013; van den Akker et al., 2020). This discrepancy might be because of women's greater likelihood of seeking help and engaging in social support networks. Such behaviors could weaken the association between

BPD and mistrust, which may account for the lack of a significant moderating effect of sex in our findings.

Education level also did not significantly moderate the effect size. Although higher educational attainment is generally associated with lower levels of mistrust (Frederiksen et al., 2016) and fewer BPD symptoms (Juurlink et al., 2022), the absence of a moderating effect might result from counterbalancing influences. Enhanced coping mechanisms associated with higher education could reduce mistrust, while increased critical thinking skills might lead to greater skepticism. These conflicting factors may cancel each other out, resulting in no significant overall effect.

The direction of the trust paradigm—whether neutral, mistrust enhancing, or trust enhancing—was not a significant moderator. All conditions showed comparable differences in mistrust between individuals with BPD and control groups. This suggests that individuals with BPD may maintain a consistent level of mistrust that is resistant to change, making it challenging to rebuild trust even in trust-enhancing situations. This finding implies that even trust-enhancing cues may not be effective in psychosocial interventions, confirming the hard to treat nature of this disorder (Fonagy et al., 2015).

The similar effects observed in neutral and mistrust-enhancing conditions imply that individuals with BPD inherently expect mistrustful outcomes, reacting to neutral situations as if they were already negative. When faced with mistrust-inducing conditions (e.g., noncooperative partners), their mistrust levels may not increase significantly because such outcomes align with their expectations. This observation aligns with findings that individuals with BPD in unpredictable environments do not readily adjust their behaviors based on advice, as they anticipate unpredictability as the norm (Fineberg et al., 2018). Therefore, situations that typically heighten mistrust do not substantially alter their mistrust levels because these outcomes are not surprising to them.

Interestingly, the comparable effect sizes across trust-enhancing, neutral, and mistrust-enhancing conditions diverge from literature documenting paradoxical trust behaviors in individuals with BPD. Previous studies (Abramov et al., 2020; Liebke et al., 2018) have suggested that mistrust-enhancing conditions can lead to increased trust, while trust-enhancing conditions may result in greater mistrust among those with BPD. In contrast, our analysis indicates that regardless of the condition, individuals with BPD exhibit higher mistrust compared to controls.

Both trustworthiness appraisals and trust behavior paradigms showed significant associations, with trustworthiness appraisals being more strongly related to the effect size. This suggests that while individuals with BPD may exhibit a mistrustful appraisal style, the interactive dynamics of behavioral tasks could potentially mitigate mistrust to some extent.

We found no significant differences between self-report and interview methods for assessing BPD. This suggests that the method of BPD assessment does not substantially influence the relationship between BPD and mistrust. One possible explanation is that, while mistrust was assessed through laboratory-based tasks, BPD symptoms were measured using either self-report or interview methods. This separation in measurement approaches likely reduces the risk of inflated effect sizes because of the absence of common method variance.

Finally, neither quality assessment scores nor publication year emerged as significant moderators in our analysis. This indicates that overall study quality was not significantly associated with the observed effect sizes. Notably, only one study (Fineberg et al., 2018) achieved an excellent quality rating, while the others did not meet this high standard.

Clinical Implications

Our findings underscore the critical importance of addressing mistrust in the treatment of BPD. Neglecting to assess and prioritize mistrust can impede the development of interpersonal trust, thereby adversely affecting epistemic trust. When new social information is perceived as nonbenign, patients may reject therapeutic interventions, significantly hindering treatment progress (Fonagy et al., 2015). Therefore, focusing on interpersonal trust is essential in clinical practice.

Several therapeutic approaches have already integrated considerations of trust issues. Mentalization-based therapy emphasizes the significance of understanding both one's own and others' mental states as foundational for building epistemic trust. A lack of this understanding can lead to deficits in epistemic trust, which is crucial for integrating new thoughts and behaviors into existing knowledge frameworks and fostering adaptable thinking patterns (Fonagy et al., 2015). Mentalization-based therapy, therefore, prioritizes establishing trustworthy therapeutic relationships as an initial step toward cultivating epistemic trust. Similarly, ST aims to reform maladaptive schemas related to interpersonal relationships (Young et al., 2003). It focuses on nurturing trustworthy and stable connections through therapist–client interactions to address mistrust and abuse schemas (Bach & Farrell, 2018). Transference-focused psychotherapy tackles the paranoid perceptions of others in BPD by striving to integrate polarized views, moving patients toward more balanced and realistic perceptions free from excessive trust or mistrust (Kernberg et al., 2008).

The smaller effect sizes observed in studies using laboratory-based and behavioral methods, compared to self-reports, suggest that individuals with BPD may be sensitive to immediate contexts, making them more receptive to interventions targeting mistrust. Additionally, the reduced effect sizes in behavioral trust paradigms compared to trustworthiness appraisal tasks indicate that the dynamic elements in behavioral tasks might alleviate inherent mistrust beliefs. Although preliminary, this insight offers a promising avenue for further research.

However, the absence of significant differences in mistrust between neutral, mistrust-enhancing, and trust-enhancing conditions contrasts with the idea that these individuals are prone to contextual influences on their mistrust levels. This finding suggests that creating a trusting therapeutic environment may be challenging, as the similar effect sizes across conditions imply that patients might not be highly receptive to trust-enhancing cues used by therapists. Therefore, our findings indicate that the complex and dynamic nature of how individuals with BPD respond to environmental influences on mistrust warrants further investigation.

Limitations of Included Evidence

Our meta-analysis relied exclusively on cross-sectional data, limiting our ability to infer temporal relationships. To assess causality and temporal dynamics more effectively, future studies should employ longitudinal designs. Another limitation pertains to the ecological validity of the included studies. Laboratory-based mistrust tasks are often perceived as abstract and artificial (Sofianos, 2022),

while some argue that self-reported perceptions of mistrust better capture real-world experiences, enhancing their relevance to clinical practice (Choi-Kain et al., 2023). Future research should incorporate stimuli involving significant others to better reflect mistrust in BPD, thereby creating designs that are more valid in real-life contexts.

Several quality issues were identified. Many studies lacked comprehensive strategies to address confounding variables. Notably, factors such as comorbidity, medication, and treatment status were not consistently controlled for, which could influence the findings. Future research should include these variables to provide a clearer understanding of the association between mistrust and BPD. Additionally, correlational studies often lacked objective diagnostic measures for BPD, and case–control studies lacked consistent criteria for participant selection and matching.

A significant limitation was the absence of CC groups in most studies, hindering our ability to determine whether mistrust is specific to BPD. Future research could include clinical comparison groups to explore if mistrust is a dysfunction specific to BPD. Incentives varied across studies, potentially affecting participant motivation and introducing heterogeneity. Future research could consider linking participation rewards to game outcomes to simulate a more realistic mistrust process in BPD. All the reviewed studies were conducted in Western Europe or the United States, which may introduce cultural bias. Broader cultural representation is needed in future research to enhance the generalizability of the findings. The Egger–Sandwich test indicated potential publication bias. Therefore, the results should be interpreted cautiously because of possible selective reporting.

Limitations of the Review Process

Potential biases may have arisen because of differing levels of reviewer involvement. One reviewer examined the entire sample, while another assessed only a 25% subset. Furthermore, our decision to include only English-language records may have excluded significant cross-cultural data, potentially affecting the generalizability of our findings. Many of the proposed moderators did not account for the heterogeneity observed in the meta-analysis results. Therefore, future analyses should consider incorporating additional variables to better explain the variability in the findings. In terms of effect size conversion, we converted correlation coefficients (r) to Hedges' g in some studies, which may have led to a loss of continuous data. This represents another limitation of our study.

Conclusion

In conclusion, this meta-analysis highlights a higher level of mistrust in individuals with BPD compared to control groups. This finding underscores the significance of mistrust as a critical concern within BPD. Our analysis supports the approach of psychosocial interventions aimed at readdressing mistrust in individuals with BPD.

References

Abramov, G., Miellet, S., Kautz, J., Grenyer, B. F. S., & Deane, F. P. (2020). The paradoxical decline and growth of trust as a function of borderline personality disorder trait count: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair. *PLoS ONE*, *15*(7), Article e0236170. https://doi.org/10.1371/journal.pone.0236170

Agrawal, H. R., Gunderson, J., Holmes, B. M., & Lyons-Ruth, K. (2004). Attachment studies with borderline patients: A review. *Harvard Review of Psychiatry*, *12*(2), 94–104. https://doi.org/10.1080/10673220490447218

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

Bach, B., & Farrell, J. M. (2018). Schemas and modes in borderline personality disorder: The mistrustful, shameful, angry, impulsive, and unhappy child. *Psychiatry Research*, *259*, 323–329. https://doi.org/10.1016/j.psychres.2017.10.039

Bateman, A. W., & Fonagy, P. (Eds.) (2012). *Handbook of mentalizing in mental health practice*. American Psychiatric Publishing.

Beck, A. T., Davis, D. D., & Freeman, A. (Eds.). (2015). *Cognitive therapy of personality disorders* (3rd ed.). The Guilford Press.

Berghuis, H., Bandell, C. C., & Krueger, R. F. (2021). Predicting dropout using DSM-5 section II personality disorders, and DSM-5 section III personality traits, in a (day)clinical sample of personality disorders. *Personality Disorders: Theory, Research, and Treatment*, *12*(4), 331–338. https://doi.org/10.1037/per0000429

Biermann, M., Schulze, A., Unterseher, F., Hamm, M., Atanasova, K., Stahlberg, D., & Lis, S. (2022). Trustworthiness judgments and Borderline Personality Disorder: An experimental study on the interplay of happiness and trustworthiness appraisals and the effects of wearing face masks during the COVID-19 pandemic in Germany. *Borderline Personality Disorder and Emotion Dysregulation*, *9*(1), Article 27. https://doi.org/10.1186/s40479-022-00193-x

Bowlby, J. (1980). *Attachment and loss*. Basic Books.

Bozzatello, P., Blua, C., Brandellero, D., Baldassarri, L., Brasso, C., Rocca, P., & Bellino, S. (2024). Gender differences in borderline personality disorder: A narrative review. *Frontiers in Psychiatry*, *15*, Article 1320546. https://doi.org/10.3389/fpsyt.2024.1320546

Brenner, P. S., & DeLamater, J. (2016). Lies, damned lies, and survey self-reports? Identity as a cause of measurement bias. *Social Psychology Quarterly*, *79*(4), 333–354. https://doi.org/10.1177/0190272516628298

Brodsky, B. S., Groves, S. A., Oquendo, M. A., Mann, J. J., & Stanley, B. (2006). Interpersonal precipitants and suicide attempts in borderline personality disorder. *Suicide and Life-Threatening Behavior*, *36*(3), 313–322. https://doi.org/10.1521/suli.2006.36.3.313

Brown, M. Z., Comtois, K. A., & Linehan, M. M. (2002). Reasons for suicide attempts and nonsuicidal self-injury in women with borderline personality disorder. *Journal of Abnormal Psychology*, *111*(1), 198–202. https://doi.org/10.1037/0021-843X.111.1.198

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

Choi-Kain, L. W., Masland, S. R., & Finch, E. F. (2023). Corrective experiences to enhance trust: Clinical wisdom from good (enough) psychiatric management. *Journal of Personality Disorders*, *37*(5), 559–579. https://doi.org/10.1521/pedi.2023.37.5.559

Clark, A. K., & Eisenstein, M. A. (2013). Interpersonal trust: An age-period-cohort analysis revisited. *Social Science Research*, *42*(2), 361–375. https://doi.org/10.1016/j.ssresearch.2012.09.006

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? *Clinical Psychology Review*, *31*(6), 965–982. https://doi.org/10.1016/j.cpr.2011.06.001

Ebert, A., Kolb, M., Heller, J., Edel, M.-A., Roser, P., & Brüne, M. (2013). Modulation of interpersonal trust in borderline personality disorder by

intranasal oxytocin and childhood trauma. *Social Neuroscience*, *8*(4), 305–313. https://doi.org/10.1080/17470919.2013.807301

Felsenheimer, A. K., Kieckhäfer, C., & Rapp, A. M. (2022). Irony detection in patients with borderline personality disorder: An experimental study examining schizotypal traits, response biases and empathy. *Borderline Personality Disorder and Emotion Dysregulation*, *9*(1), Article 24. https://doi.org/10.1186/s40479-022-00194-w

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). Visual representations of meta-analyses of multiple outcomes: Extensions to forest plots, funnel plots, and caterpillar plots. *Methodology*, *16*(4), 299–315. https://doi.org/10.5964/meth.4013

Fertuck, E. A., Fischer, S. A., & Melara, R. D. (2023). Atypical neural plasticity and behavioral effects of trustworthiness learning in borderline personality disorder features. *Journal of Personality Disorders*, *37*(5), 542–558. https://doi.org/10.1521/pedi.2023.37.5.542

Fertuck, E. A., Grinband, J., Mann, J. J., Hirsch, J., Ochsner, K., Pilkonis, P., Erbe, J., & Stanley, B. (2019). Trustworthiness appraisal deficits in borderline personality disorder are associated with prefrontal cortex, not amygdala, impairment. *NeuroImage: Clinical*, *21*, Article 101616. https://doi.org/10.1016/j.nicl.2018.101616

Fineberg, S. K., Leavitt, J., Stahl, D. S., Kronemer, S., Landry, C. D., Alexander-Bloch, A., Hunt, L. T., & Corlett, P. R. (2018). Differential valuation and learning from social and nonsocial cues in borderline personality disorder. *Biological Psychiatry*, *84*(11), 838–845. https://doi.org/10.1016/j.biopsych.2018.05.020

Fisher, Z., Tipton, E., & Zhipeng, H. (2023). *robumeta: Robust variance meta-regression* (R package Version 2.1) [Computer software]. https://CRAN.R-project.org/package=robumeta

Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *Journal of Personality Disorders*, *29*(5), 575–609. https://doi.org/10.1521/pedi.2015.29.5.575

Frederiksen, M., Larsen, C. A., & Lolle, H. (2016). Education and trust: Exploring the association across social relationships and nations. *Acta Sociologica*, *59*(4), 293–308. https://doi.org/10.1177/0001699316658936

Freeman, D., & Garety, P. A. (2000). Comments on the content of persecutory delusions: Does the definition need clarification? *British Journal of Clinical Psychology*, *39*(4), 407–414. https://doi.org/10.1348/014466500163400

Frías, Á, Palma, C., Solves, L., Martínez, B., & Salvador, A. (2017). Differential symptomatology and functioning in borderline personality disorder across age groups. *Psychiatry Research*, *258*, 44–50. https://doi.org/10.1016/j.psychres.2017.09.081

Grant, B. F., Chou, S. P., Goldstein, R. B., Huang, B., Stinson, F. S., Saha, T. D., Smith, S. M., Dawson, D. A., Pulay, A. J., Pickering, R. P., & Ruan, W. J. (2008). Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: Results from the Wave 2 National Epidemiologic Survey on alcohol and related conditions. *The Journal of Clinical Psychiatry*, *69*(4), 533–545. https://doi.org/10.4088/JCP.v69n0404

Gunderson, J. G. (2007). Disturbed relationships as a phenotype for borderline personality disorder. *American Journal of Psychiatry*, *164*(11), 1637–1640. https://doi.org/10.1176/appi.ajp.2007.07071125

Gunderson, J. G., Stout, R. L., McGlashan, T. H., Shea, M. T., Morey, L. C., Grilo, C. M., Zanarini, M. C., Yen, S., Markowitz, J. C., Sanislow, C., Ansell, E., Pinto, A., & Skodol, A. E. (2011). Ten-year course of borderline personality disorder: Psychopathology and function from the collaborative longitudinal personality disorders study. *Archives of General Psychiatry*, *68*(8), 827–837. https://doi.org/10.1001/archgenpsychiatry.2011.37

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. (2021). *Doing meta-analysis with R: A hands-on guide* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003107347

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Howard, K. P., Lazarus, S. A., & Cheavens, J. S. (2022). A longitudinal examination of the reciprocal relationship between borderline personality features and interpersonal relationship quality. *Personality Disorders: Theory, Research, and Treatment*, *13*(1), 3–11. https://doi.org/10.1037/per0000484

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, *75*(3), 334–349. https://doi.org/10.1037/0021-9010.75.3.334

Juurlink, T. T., Betts, J. K., Nicol, K., Lamers, F., Beekman, A. T. F., Cotton, S. M., & Chanen, A. M. (2022). Characteristics and predictors of educational and occupational disengagement among outpatient youth with borderline personality disorder. *Journal of Personality Disorders*, *36*(1), 116–128. https://doi.org/10.1521/pedi_2021_35_534

Kaplan, J. (2024). *fastDummies: Fast creation of dummy (binary) columns and rows from categorical variables* (R package Version 1.7.4) [Computer software]. https://CRAN.R-project.org/package=fastDummies

Kernberg, O. F. (1967). Borderline personality organization. *Journal of the American Psychoanalytic Association*, *15*(3), 641–685. https://doi.org/10.1177/000306516701500309

Kernberg, O. F., Yeomans, F. E., Clarkin, J. F., & Levy, K. N. (2008). Transference focused psychotherapy: Overview and update. *The International Journal of Psychoanalysis*, *89*(3), 601–620. https://doi.org/10.1111/j.1745-8315.2008.00046.x

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Kurt, Y., Walker, M., Luyten, P., & Fonagy, P. (2025). *Behavioral and laboratory-assessed mistrust in borderline personality disorder: A meta-analytic review*. https://accounts.osf.io/login?service=https%3A%2F%2Fosf.io%2Ft6syu%2F

Lazarus, S. A., Cheavens, J. S., Festa, F., & Rosenthal, M. Z. (2014). Interpersonal functioning in borderline personality disorder: A systematic review of behavioral and laboratory-based assessments. *Clinical Psychology Review*, *34*(3), 193–205. https://doi.org/10.1016/j.cpr.2014.01.007

Liebke, L., Koppe, G., Bungert, M., Thome, J., Hauschild, S., Defiebre, N., Izurieta Hidalgo, N. A., Schmahl, C., Bohus, M., & Lis, S. (2018). Difficulties with being socially accepted: An experimental study in borderline personality disorder. *Journal of Abnormal Psychology*, *127*(7), 670–682. https://doi.org/10.1037/abn0000373

Masland, S. R., Schnell, S. E., & Shah, T. V. (2020). Trust beliefs, biases, and behaviors in borderline personality disorder: Empirical findings and relevance to epistemic trust. *Current Behavioral Neuroscience Reports*, *7*(4), 239–249. https://doi.org/10.1007/s40473-020-00220-7

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of *r* and *d*. *Psychological Methods*, *11*(4), 386–401. https://doi.org/10.1037/1082-989X.11.4.386

Miano, A., Lustig, S., Meyerding, L., & Barnow, S. (2023). Accuracy and bias in facial trustworthiness appraisals in borderline personality disorder. *Journal of Personality Disorders*, *37*(5), 525–541. https://doi.org/10.1521/pedi.2023.37.5.525

Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Qureshi, R., Mattis, P., Lisy, K., & Mu, P.-F. (2020). Systematic reviews of etiology and risk. In E. Aromataris & Z. Munn (Eds.), *JBI manual for evidence synthesis* (pp. 252–311). JBI. https://synthesismanual.jbi.global

Mullins-Sweatt, S. N., Edmundson, M., Sauer-Zavala, S., Lynam, D. R., Miller, J. D., & Widiger, T. A. (2012). Five-factor measure of borderline personality traits. *Journal of Personality Assessment*, *94*(5), 475–487. https://doi.org/10.1080/00223891.2012.672504

Nicol, K., Pope, M., Sprengelmeyer, R., Young, A. W., & Hall, J. (2013). Social judgement in borderline personality disorder. *PLoS ONE*, *8*(11), Article e73440. https://doi.org/10.1371/journal.pone.0073440

Niedtfeld, I., & Kroneisen, M. (2020). Impaired memory for cooperative interaction partners in borderline personality disorder. *Borderline*

Personality Disorder and Emotion Dysregulation, 7(1), Article 22. https://doi.org/10.1186/s40479-020-00137-3

Oldham, J. M. (2006). Borderline personality disorder and suicidality. *The American Journal of Psychiatry*, 163(1), 20–26. https://doi.org/10.1176/appi.ajp.163.1.20

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., & Chou, R. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71. https://doi.org/10.1136/bmj.n71

Preti, E., Richetin, J., Poggi, A., & Fertuck, E. (2023). A model of trust processes in borderline personality disorder: A systematic review. *Current Psychiatry Reports*, 25(11), 555–567. https://doi.org/10.1007/s11920-023-01468-y

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Roberts, I. D., Krajbich, I., Cheavens, J. S., Campo, J. V., & Way, B. M. (2018). Acetaminophen reduces distrust in individuals with borderline personality disorder features. *Clinical Psychological Science*, 6(1), 145–154. https://doi.org/10.1177/2167702617731374

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, 26(2), 141–160. https://doi.org/10.1037/met0000300

Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and *DSM-IV-TR* personality disorders: A facet level analysis. *Clinical Psychology Review*, 28(8), 1326–1342. https://doi.org/10.1016/j.cpr.2008.07.002

Scheunemann, J., Jelinek, L., Biedermann, S. V., Lipp, M., Yassari, A. H., Kühn, S., Gallinat, J., & Moritz, S. (2023). Can you trust this source? Advice taking in borderline personality disorder. *European Archives of Psychiatry and Clinical Neuroscience*, 273(4), 875–885. https://doi.org/10.1007/s00406-022-01539-w

Shapiro-Thompson, R., Shah, T. V., Yi, C., Jackson, N., Diaz, D. T., & Fineberg, S. K. (2023). Modulation of trust in borderline personality disorder by script-based imaginal exposure to betrayal. *Journal of Personality Disorders*, 37(5), 508–524. https://doi.org/10.1521/pedi.2023.37.5.508

Sharp, C. (2016). Current trends in BPD research as indicative of a broader sea-change in psychiatric nosology. *Personality Disorders: Theory, Research, and Treatment*, 7(4), 334–343. https://doi.org/10.1037/per0000199

Sharp, C., & Vanwoerden, S. (2015). Hypermentalizing in borderline personality disorder: A model and data. *Journal of Infant, Child, and Adolescent Psychotherapy*, 14(1), 33–45. https://doi.org/10.1080/15289168.2015.1004890

Sharp, C., Venta, A., & McLaren, V. (2023). Interpersonal trust in adolescents with borderline personality disorder: Comparisons with healthy and psychiatric controls. *Journal of Personality Disorders*, 37(5), 475–489. https://doi.org/10.1521/pedi.2023.37.5.475

Skodol, A. E., Pagano, M. E., Bender, D. S., Shea, M. T., Gunderson, J. G., Yen, S., Stout, R. L., Morey, L. C., Sanislow, C. A., Grilo, C. M., Zanarini, M. C., & McGlashan, T. H. (2005). Stability of functional impairment in patients with schizotypal, borderline, avoidant, or obsessive–compulsive personality disorder over two years. *Psychological Medicine*, 35(3), 443–451. https://doi.org/10.1017/S003329170400354X

Sofianos, A. (2022). Self-reported & revealed trust: Experimental evidence. *Journal of Economic Psychology*, 88, Article 102451. https://doi.org/10.1016/j.joep.2021.102451

Thimm, J. C., & Chang, M. (2022). Early maladaptive schemas and mental disorders in adulthood: A systematic review and meta-analysis. *International Journal of Cognitive Therapy*, 15(4), 371–413. https://doi.org/10.1007/s41811-022-00149-7

Unoka, Z., Seres, I., Áspán, N., Bódi, N., & Kéri, S. (2009). Trust game reveals restricted interpersonal transactions in patients with borderline personality disorder. *Journal of Personality Disorders*, 23(4), 399–409. https://doi.org/10.1521/pedi.2009.23.4.399

van den Akker, O. R., van Assen, M. A. L. M., van Vugt, M., & Wicherts, J. M. (2020). Sex differences in trust and trustworthiness: A meta-analysis of the trust game and the gift-exchange game. *Journal of Economic Psychology*, 81, Article 102329. https://doi.org/10.1016/j.joep.2020.102329

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, 47(4), 1274–1294. https://doi.org/10.3758/s13428-014-0527-2

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Watters, C. A., Bagby, R. M., & Sellbom, M. (2019). Meta-analysis to derive an empirically based set of personality facet criteria for the alternative *DSM-5* model for personality disorders. *Personality Disorders: Theory, Research, and Treatment*, 10(2), 97–104. https://doi.org/10.1037/per0000307

Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema therapy: A practitioner's guide*. Guilford Press.

Yuan, Y., Lee, H., Eack, S. M., & Newhill, C. E. (2023). A systematic review of the association between early childhood trauma and borderline personality disorder. *Journal of Personality Disorders*, 37(1), 16–35. https://doi.org/10.1521/pedi.2023.37.1.16