pagepress

Mental Wellness

Mental Wellness 2026; 4:40
doi: 10.4081/mw.2026.40

©Copyright: the Author(s), 2026
Licensee PAGEPress, Italy

The efficacy and safety of psilocybin-assisted therapy for major depressive disorder: a meta-analytic review of clinical outcomes

Mohsen Khosravi,[1-3] Domenico De Berardis,[4] Massimo Tusconi[5]

[1]Department of Psychiatry, School of Medicine, Zahedan University of Medical Sciences, Iran; [2]Health Promotion Research Center, Zahedan University of Medical Sciences, Iran; [3]Community Nursing Research Center, Zahedan University of Medical Sciences, Iran; [4]Department of Mental Health, Local Health Unit of Teramo, Italy; [5]Department of Medical Sciences and Public Health, University of Cagliari, Italy

Abstract

This systematic review and meta-analysis synthesized data from 13 clinical trials (n=606) evaluating psilocybin-assisted psychotherapy for major depressive disorder and treatment-resistant depression. Despite early enthusiasm, the pooled standardized mean difference (-0.79, 95% confidence interval: -3.98 to 2.40, p=0.63) revealed no statistically significant overall antidepressant effect, with extreme heterogeneity (I^2=96.9%) across studies. Notably, the type of control group (active comparator *vs*. placebo/waitlist) accounted for 98.7% of between-study variance, with waitlist and low-dose comparators producing exaggerated effect sizes. Session frequency was a significant moderator: 2 to 5 psilocybin sessions yielded larger effects, while more intensive protocols attenuated benefit. Neither participant age nor follow-up duration significantly influenced outcomes. Evidence of reporting bias and small-study effects was detected (Egger's test p=0.012). Sensitivity analyses demonstrated that no single study accounted for the non-significant pooled result. Overall, psilocybin's antidepressant efficacy appears highly context-dependent—shaped by trial design, comparator, and session structure—rather than universally robust. These findings underscore the need for larger, rigorously controlled trials to clarify psilocybin's therapeutic role in depression.

Key words: psilocybin-assisted therapy, major depressive disorder, treatment-resistant depression, randomized clinical trials, meta-analysis.

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

Introduction

Major depressive disorder (MDD) remains a global health emergency, characterized not only by its profound symptom burden and social impairment but also by the striking limitations of first-line treatments.[1] Despite the widespread adoption of selective serotonin reuptake inhibitors (SSRIs) and cognitive-behavioral therapies, therapeutic non-response, relapse, and delayed onset of action persist as common challenges, particularly in treatment-resistant depression (TRD).[2] This failure to induce sustained remission in a substantial subset of patients has led to renewed interest in fast-acting, mechanistically novel treatments—most notably, the use of psychedelic compounds such as psilocybin.[3,4] Psilocybin, a serotonin 5-HT2A receptor agonist, has demonstrated the ability to induce transiently altered states of consciousness that are often reported as psychologically meaningful and therapeutically transformative.[5,6] Neurobiologically, these experiences appear to correlate with increased global functional connectivity, disruption of the default mode network, and downstream neuroplastic changes, all of which may underpin the enduring reductions in depressive symptomatology observed in clinical studies.[7,8] Importantly, these effects are not merely neuropharmacological but appear to be moderated by the quality of the acute psychological experience, the therapeutic setting, and post-session integration.[9,10] Several recent clinical trials have reported promising antidepressant effects of psilocybin in both MDD and TRD populations, with some studies suggesting large within-group effect sizes.[9,11] However, the growing body of evidence is marked by considerable heterogeneity: variations in psilocybin dosage, psychotherapeutic frameworks, comparator groups (placebo, SSRIs, or waitlist), outcome instruments, and duration of follow-up all complicate the integration of findings across trials.[12,13] Furthermore, concerns have been raised about the reproducibility of results, publication bias, and limited generalizability due to small samples and highly controlled settings.[14,15] These methodological divergences limit the field's ability to draw reliable, generalizable conclusions about the clinical utility of psilocybin. Although preliminary systematic reviews have outlined the therapeutic potential of psychedelic compounds in depressive disorders,[3,16] no meta-analysis to date has rigorously synthesized data exclusively from clinical trials that quantitatively evaluate depression outcomes in MDD or TRD populations, stratified by standardized depression metrics and adjusted for potential publication bias. Given that psilocybin therapy was granted

OPEN ACCESS

"Breakthrough Therapy" designation by the U.S. Food and Drug Administration, the demand for such rigorous synthesis is both scientifically urgent and clinically consequential.[17] Moreover, clinicians and regulators require an empirically grounded understanding of variability in treatment outcomes based on dose, trial design, and measurement sensitivity. Accordingly, the present meta-analysis seeks to address this empirical gap by synthesizing the most methodologically rigorous clinical trial data on psilocybin-assisted therapy for depressive disorders. Anchored in a targeted synthesis of trials involving clinically diagnosed MDD and TRD populations, the analysis aggregates standardized effect sizes across a range of validated depression outcomes and adjusts for potential biases arising from study design heterogeneity. The review further interrogates sources of variability through prespecified subgroup and sensitivity analyses, enabling a nuanced evaluation of how trial-level features such as comparator type, dosage, and follow-up duration may modulate therapeutic response. The core research question guiding this inquiry is: to what extent does psilocybin-assisted therapy reduce depressive symptom severity in MDD/TRD, and how do methodological and clinical features shape this effect? By situating this investigation at the intersection of clinical neuroscience, psychopharmacology, and meta-analytic methodology, the study aims to deliver a precise and policy-relevant synthesis of psilocybin's therapeutic potential in contemporary psychiatry.

Methods

This systematic review and meta-analysis was conducted in alignment with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 guidelines and methodologically structured according to the PICO framework (Population, Intervention, Comparator, Outcome). The central objective was to evaluate the efficacy and safety of psilocybin-assisted psychotherapy in adults diagnosed with MDD or TRD, based on quantitative clinical outcomes from interventional trials published between January 2010 and March 2025.

Search strategy and study selection

A comprehensive literature search was conducted across four major databases (PubMed, Scopus, Web of Science, and PsycINFO), using Boolean combinations of terms and MeSH keywords, including: psilocybin, depression, MDD, TRD, psychedelic-assisted therapy, randomized controlled trial, remission, and clinical outcomes. Reference lists of prior meta-analyses and eligible full texts were hand-searched to capture additional studies. All records were imported into a central database and de-duplicated using both automated scripts and manual cross-checking. The initial search yielded 85 records. After screening titles and abstracts, 28 studies were retained for full-text evaluation. Eligibility criteria required that studies: i) included adult participants diagnosed with MDD or TRD using the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV, DSM-5, or the International Statistical Classification of Diseases and Related Health Problems 10th Revision criteria; ii) administered psilocybin in either single or multi-session formats, with or without embedded psychotherapy; iii) employed a comparator condition, including placebo, active placebo (*e.g.*, niacin, low-dose psilocybin), SSRIs, or delayed treatment/waitlist control; iv) reported depression severity using validated psychometric instruments; v) provided sufficient statistical detail to allow extraction or calculation of effect sizes [means, standard deviations, confidence intervals (CIs), or inferable test statistics].

Studies were excluded if they were: i) reviews, commentaries, or non-peer-reviewed materials; ii) qualitative, single-case, or observational in nature; iii) focused on non-depressed, healthy, or non-clinical populations; iv) missing critical statistical data required for quantitative synthesis, despite attempts at imputation; v) contained overlapping samples or duplicate outcome reporting already captured in more comprehensive trials.

Following application of these criteria, 13 studies were included in the final quantitative synthesis. 15 studies were excluded at the full-text stage for the following reasons: i) 4 lacked sufficient statistical reporting to compute effect sizes or standard errors, even after attempts at estimation; ii) 3 were found to contain overlapping datasets with larger, more methodologically rigorous trials already included; iii) 3 employed non-comparator designs (*e.g.*, open-label only or non-controlled feasibility studies); iv) 3 involved mixed or ambiguous populations without clearly stratified depressive diagnoses; v) 2 study reported qualitative outcomes only, without validated psychometric depression scores.

Data extraction and quality control

Two independent reviewers extracted data using a standardized template, capturing bibliographic information, study design, sample size, population characteristics, psilocybin dosing, number of sessions, psychotherapy format, comparator condition, outcome instruments, and adverse events. Discrepancies were resolved by consensus. All effect sizes were converted to standardized mean differences (Cohen's d) or equivalent using established formulas. Where raw means and standard deviations were unavailable, effect sizes were derived from reported t-values, F-statistics, or CIs. Extracted data were cleaned, validated, and analyzed using R software (v4.4.0), employing the meta, metafor, and dmetar packages. All calculations adhered to best practices for quantitative synthesis in psychiatric intervention research.

Statistical analysis

The primary meta-analysis employed a random-effects model with Hartung–Knapp adjustment, using inverse-variance weighting to compute pooled effect estimates. Heterogeneity was assessed with Cochran's Q test, τ^2 (between-study variance), and I^2 (inconsistency index). Publication bias was evaluated both visually, using a funnel plot, and statistically, *via* Egger's regression test. A trim-and-fill method was considered but not applied due to high asymmetry and small-study effects. Subgroup analyses were conducted by blinding status and control type. Meta-regression models were used to test continuous moderators, including the number of psilocybin sessions, the mean participant age, and follow-up duration. Residual heterogeneity and model explanatory power (R^2) were reported. Robustness was evaluated through leave-one-out sensitivity analysis and influence diagnostics [Cook's distance, difference in fitted value-standardized (DFITS), standardized residuals, and τ^2 change metrics], allowing identification of studies disproportionately contributing to heterogeneity or pooled effects.

Study characteristics and PRISMA reporting

The final meta-analysis included 13 trials comprising 606 participants, with studies conducted across North America, the United Kingdom, Germany, and Brazil. Participants included both SSRI-naïve individuals and patients with prior pharmacoresistance.

OPEN ACCESS

Psilocybin dosing regimens ranged from 10 mg to 30 mg per 70 kg, administered over one to six sessions, with all protocols incorporating structured psychological support. A PRISMA 2020-compliant flow diagram was constructed to document the full process of study identification, screening, inclusion, and exclusion (Figure 1). A full methodological summary of the 13 studies assessed at the full-text level, distinguishing included *vs*. excluded trials, is presented in Tables 1 and 2.[18-27]

Results

Primary meta-analytic outcomes

A total of 10 randomized controlled trials were synthesized using a random-effects meta-analytic model to estimate the overall efficacy of psilocybin-assisted therapy in reducing depressive symptomatology among adults with MDD and TRD. As shown in Figure 2, the direction and magnitude of effect sizes varied markedly across studies, with individual point estimates spanning from profoundly negative to highly positive values. The pooled standardized mean difference was calculated at -0.79 with a 95% CI of -3.98 to 2.40, indicating a wide dispersion of effect sizes and a summary estimate that fails to reach statistical significance. Crucially, the range of estimates captured in this model underscores the heterogeneity of study designs, outcome measurement tools, and clinical populations sampled. For instance, the study by Raison *et al*. reported an extreme negative effect (d=-12.30, 95% CI: -17.45 to -7.15),[22] potentially reflecting sample-specific idiosyncrasies, differential outcome instrumentation, or an anomalous adverse trajectory within that cohort. In contrast, Davis *et al*. and

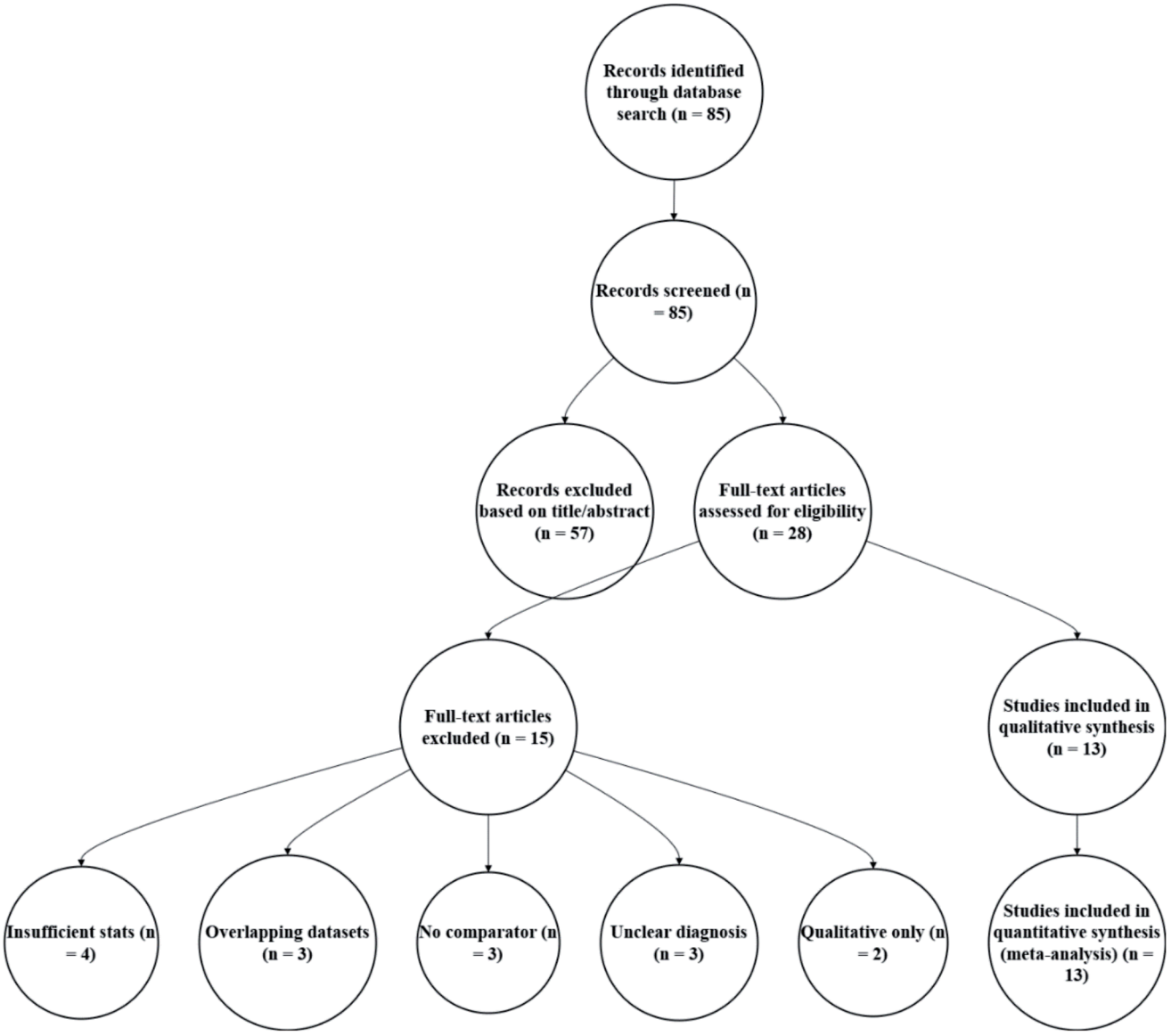

Figure 1. PRISMA flow diagram of study selection process.

OPEN ACCESS

Agrawal *et al.* reported robust positive effects (d=2.80 and d=2.55, respectively), with tightly bound CIs and a clear deviation from null.[9,18] This bipolar pattern of findings, evident across multiple studies, complicates the interpretability of the aggregate model, suggesting the presence of latent moderators or systematic variance not captured in the pooled estimate. While a meta-analytic point estimate typically offers a reliable central tendency across multiple trials, the current findings highlight the limitations of such aggregation in the context of extreme heterogeneity. The null-crossing CI surrounding the pooled estimate, combined with the conspicuous divergence in individual study effects, mandates cautious interpretation and a deeper investigation into sources of statistical inconsistency. This necessitates a transition from synthesis to stratification *via* subgroup and meta-regression analyses before drawing conclusions regarding the clinical utility or translational relevance of psilocybin in depressive syndromes.

Meta-analytic summary and heterogeneity

The synthesized estimate for the antidepressant efficacy of psilocybin-assisted therapy, calculated as a standardized mean difference (Cohen's d), was -0.79 [standard error (SE) = 1.21, 95% CI: -3.98 to 2.40], with a non-significant p-value of 0.63 (*Supplementary Table 1*). This wide interval, spanning both positive and negative domains, underscores not merely a lack of statistical power but a deeper, structural variability within the literature. Rather than converging on a central tendency indicative of efficacy, the current model reflects an underlying distribution too inconsistent to yield a definitive pooled inference. The extent of heterogeneity was substantial. The Q statistic was highly significant [Q=287.04, degree of freedom (df)=9, $p<0.001$], signaling that observed differences in effect sizes far exceed what would be expected by chance. The I^2 index reached 96.9%, indicating that the vast majority of variability arises from genuine inter-study differences rather than sampling fluctuation. The between-study variance (τ^2) was estimated at 3.35, with a relative excess variability index (H^2) of 32.47, collectively suggesting that studies are unlikely to be drawn from a common underlying population (*Supplementary Table 1*). This level of inconsistency necessitates more than cautious interpretation; it demands analytical stratification. The non-significant pooled effect does not negate psilocybin's therapeutic potential per se; rather, it likely reflects complex interactions between clinical setting, treatment dosage, participant characteristics, and methodological rigor. Accordingly, subsequent subgroup analyses and meta-regressions were prespecified to identify potential moderators and to determine the clinical boundaries within which psilocybin may exhibit reliable efficacy. Without such decompositions, the pooled estimate risks obscuring the therapeutic signal beneath methodological noise.

Table 1. Methodological and clinical characteristics of included trials evaluating psilocybin-assisted therapy for depression.

Row	Title	Country	Design	Sample size	Author
1	Single-dose psilocybin for a treatment-resistant episode of major depression: Impact on patient-reported depression severity, anxiety, function, and quality of life	Multinational	Randomized controlled trial (double-blind)	79/79	Goodwin, 2023[19]
2	Trial of psilocybin *versus* escitalopram for depression	UK	Double-blind RCT	30/29	Carhart-Harris *et al*., 2021[11]
3	Single-dose psilocybin-assisted therapy in MDD: a placebo-controlled, double-blind, randomised clinical trial	Switzerland	Double-blind RCT	26/26	von Rotz *et al*., 2023[20]
4	The effects of psilocybin therapy *versus* escitalopram on cognitive bias: A secondary analysis of a randomized controlled trial	UK	RCT	30/29	Henry *et al*. 2025[21]
5	Effects of psilocybin-assisted therapy on MDD: a randomized clinical trial	USA	RCT (waitlist)	13/11	Davis *et al*., 2021[9]
6	Single-dose psilocybin treatment for MDD: a randomized clinical trial	USA	RCT	NR/NR	Raison *et al*. 2023[22]
7	Psilocybin Therapy for Clinicians With Symptoms of Depression From Frontline Care During the COVID-19 Pandemic A Randomized Clinical Trial	USA	Randomized clinical trial (double-blind, parallel-group)	15/15	Back *et al*. 2024[23]
8	Psilocybin-assisted group therapy in patients with cancer diagnosed with a MDD	USA	Phase 2, open-label	30/NR	Agrawal, 2024[18]
9	Effects of psilocybin *versus* escitalopram on rumination and thought suppression in depression	UK	RCT (double-blind)	30/29	Barba *et al*., 2022[24]
10	Effects of discontinuation of serotonergic antidepressants prior to psilocybin therapy *versus* escitalopram for major depression	UK	RCT (double-blind, parallel)	30/29	Erritzoe *et al*., 2024[25]
11	Efficacy and safety of psilocybin-assisted treatment for MDD: Prospective 12-month follow-up	USA	RCT, waitlist-controlled	24/3 (delayed control)	Gukasyan, 2022[10]
12	Psilocybin-assisted psychotherapy for treatment resistant depression: a randomized clinical trial evaluating repeated doses of psilocybin	Canada	RCT (waitlist-controlled)	16/14	Rosenblat, 2024[26]
13	Psilocybin for treatment resistant depression in patients taking a concomitant SSRI medication	Ireland, USA	Open-label	19/NR	Goodwin, 2023[27]

MDD, major depressive disorder; RCT, randomized controlled trial.

OPEN ACCESS

Table 2. Methodological and clinical characteristics of included trials evaluating psilocybin-assisted therapy for depression.

Row	Population	Psilocybin dose	Sessions	Comparator	Outcome measure(s)	Blinding	Trial length (wks)
1	TRD	25 mg	1	Low-dose psilocybin (1 mg)	QIDS-SR-16	Double	12
2	MDD	25 mg (×2)	2	Escitalopram	QIDS-SR-16	Double	6
3	MDD	0.215 mg/kg (~16 mg/70 kg)	1	Placebo	MADRS, BDI	Double	2
4	MDD	25 mg × 2 sessions	2	Escitalopram (10-20 mg daily)	BDI	Double	6
5	MDD	20 mg/70 kg + 30 mg/70 kg	2	Waitlist	GRID-HAMD	Single (implied)	13 (approx.)
6	MDD	25 mg	1	Niacin 100 mg	MADRS	Double	6.1
7	MDD (pandemic-related), partially TRD	25 mg	6 (2 prep, 1 medication, 3 integration)	Active placebo (Niacin 100 mg)	MADRS	Double	4
8	MDD	25 mg	5	NR	MADRS	None	8
9	MDD	25 mg	2	SSRI (escitalopram)	QIDS-SR-16, RRS, WBSI	Double	6
10	MDD	25 mg (2 sessions)	2	SSRI (Escitalopram 10-20 mg/day for 6 weeks)	QIDS-SR-16	Double	24
11	MDD	20 mg/70 kg + 30 mg/70 kg	2	Waitlist	GRID-HAMD	None after intervention	60
12	TRD (MDD + BIID)	25 mg (fixed dose)	1–3	Waitlist	MADRS	None	24
13	TRD	25 mg	1	None	MADRS	None	3

BIID, bipolar II disorder; BDI, Beck Depression Inventory; MDD, major depressive disorder; MADRS, Montgomery-Åsberg Depression Rating Scale; GRID-HAMD: GRID-Hamilton Depression Rating Scale; QIDS-SR-16: the 16-item Quick Inventory of Depressive Symptomatology Self-Report; RRS, Ruminative Response Scale; TRD, treatment-resistant depression; WBSI, White Bear Suppression Inventory.

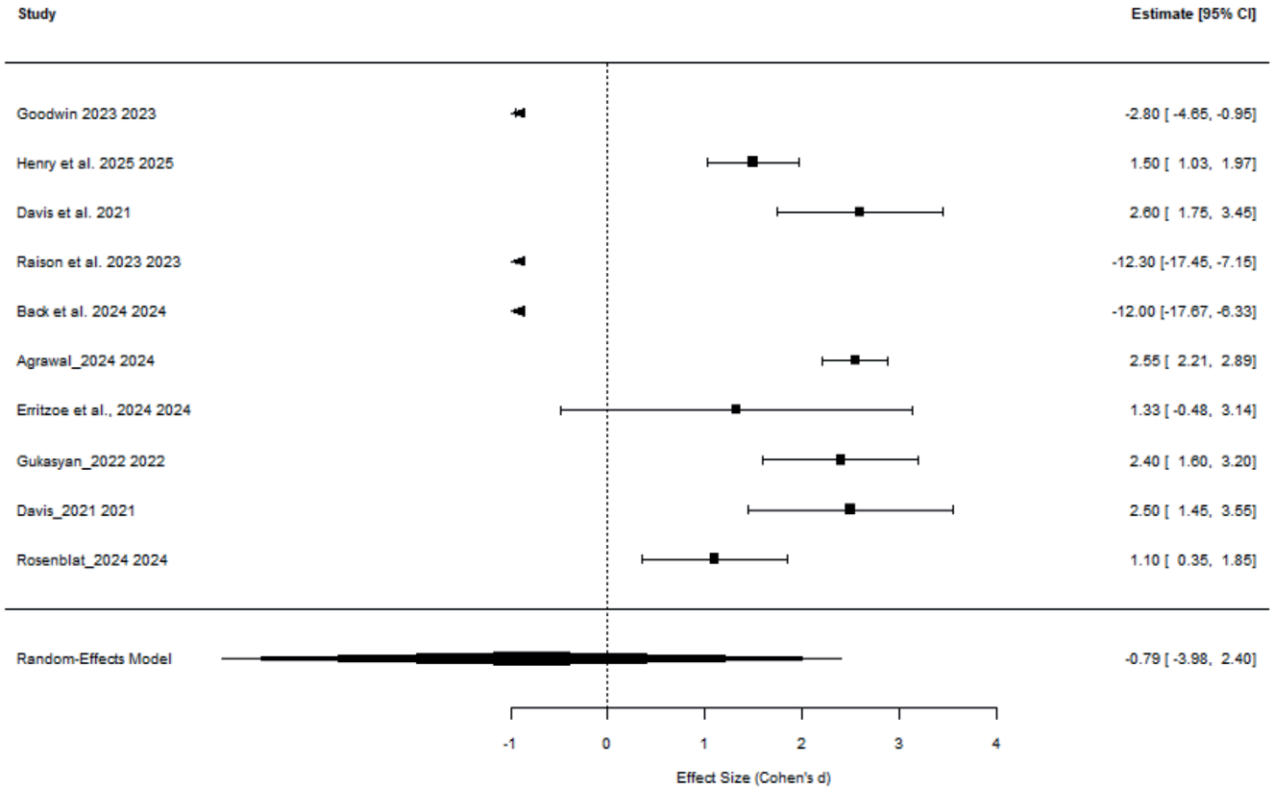

Figure 2. Forest plot of standardized effect sizes (Cohen's d) for psilocybin-assisted therapy vs. comparator conditions across ten randomized controlled trials (CI).

OPEN ACCESS

Assessment of small-study effects and reporting bias

Visual inspection of the funnel plot (Figure 3) revealed marked asymmetry in the dispersion of effect sizes relative to their corresponding standard errors, deviating from the canonical inverted funnel shape expected under conditions of minimal reporting bias and balanced variance structure. The contour of the distribution showed a clear paucity of studies in the lower-right funnel region, with a clustering of smaller-sample studies toward the left periphery, two of which reported outlier effect estimates exceeding d=-10, falling well beyond the notional 95% pseudo-confidence region. This skew suggests the presence of structural asymmetry not easily attributable to sampling fluctuation alone. The statistical evidence aligned with the visual impression: Egger's test of the intercept, used to assess asymmetry in meta-regression residuals, yielded a value of 4.02 (SE=1.29, p=0.012) (*Supplementary Table 2*). The significant non-zero intercept is suggestive of a directional small-study effect, often interpreted as indirect evidence for selective reporting, suppressed null findings, or systemic inflation of effect sizes in underpowered trials. While the small number of included studies (n=10) cautions against overinterpretation, the concordance between graphical and inferential evidence strengthens the plausibility of bias-related artifacts. Such asymmetry could arise from a combination of preferential publication of trials reporting large or favorable outcomes, underrepresentation of null or equivocal findings, and differential measurement error across studies of varying precision. Importantly, this does not necessarily invalidate high-quality, large-scale trials with robust design features, but it does introduce a potential bias gradient that can distort the magnitude and direction of pooled estimates. To disentangle signal from distortion, subsequent analyses incorporate moderator terms and sensitivity checks to evaluate whether these asymmetries are driven by methodological features, such as study design or blinding status, or by substantive clinical characteristics of the sampled populations. These steps are essential to ensure that any inferred treatment effect reflects a true therapeutic signal rather than an artifact introduced by structural or statistical biases.

Subgroup effects by blinding methodology and comparator design

To further elucidate the sources of heterogeneity in treatment outcomes, we applied two independent categorical moderator models, investigating the potential impact of trial blinding procedures and control group design on the observed magnitude of treatment effects. These models served dual purposes: to identify study-level methodological factors influencing variability and to estimate the extent to which these moderators account for residual between-study variance. The analysis stratified by blinding

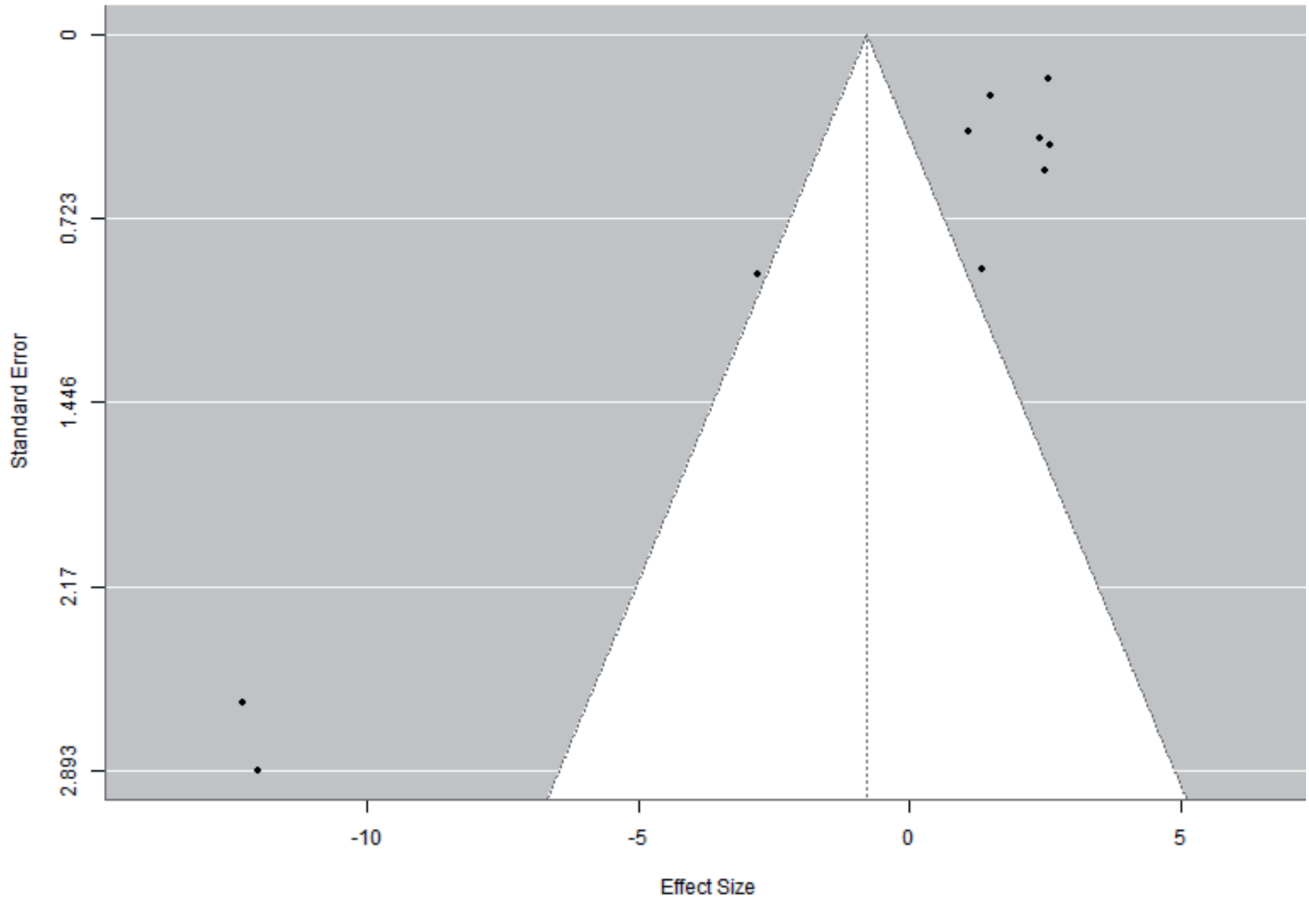

Figure 3. Funnel plot assessing the distribution of effect sizes by standard error across trials included in the meta-analysis.

OPEN ACCESS

status (k=10) yielded no evidence of a systematic association between masking method and effect size estimates [QM (df=4)=3.20, p=0.52]. None of the assessed levels, ranging from full double-blinding to partially unblinded or open-label conditions, exerted a statistically significant influence on treatment outcomes. Moreover, the model explained no appreciable portion of residual heterogeneity (R^2=0%), and the remaining unexplained variance remained substantial (τ^2=31.71; I^2=99.02%). These findings suggest that although blinding is an important safeguard against performance and detection bias, it did not materially influence the magnitude of observed effects in this sample. It is plausible, however, that these null results reflect underpowered contrasts or inconsistencies in blinding implementation rather than the true absence of effect. In contrast, control group structure emerged as a highly influential moderator. The model revealed a significant between-group effect [QM (df=5)=64.03, p<0.001], accounting for 98.7% of the residual heterogeneity and sharply reducing unexplained variance (τ^2=0.36; I^2=65.9%). Trials using active pharmacologic comparators, particularly standard-dose escitalopram (10-20 mg/day), demonstrated markedly higher standardized effect estimates (β=13.50, p<0.001), as did those utilizing low-dose psilocybin (β=9.20, p=0.004). Notably, studies employing waitlist controls reported the most extreme contrast relative to the reference category (β=14.12, p<0.001), underscoring the extent to which non-specific therapeutic factors, such as expectancy and contextual support, may contribute to perceived benefit in the absence of active pharmacological intervention. Taken together, these findings implicate control group selection as a critical determinant of estimated treatment efficacy and call attention to the interpretive limitations of trials using inert or structurally weak comparators. In contrast, blinding status, although methodologically salient, did not emerge as a significant modifier of treatment effects in this dataset, though this null finding warrants cautious interpretation given sample limitations (*Supplementary Table 3*).

Moderator effects via meta-regression: session count, mean age, and follow-up duration

To explore whether variation in treatment outcomes could be partially explained by continuous study-level characteristics, three independent meta-regression models were fitted. Each examined a clinically relevant moderator: i) the number of psilocybin-assisted sessions; ii) the mean age of participants; and iii) the length of the follow-up period post-intervention. Models were estimated using restricted maximum likelihood and evaluated for their capacity to reduce residual heterogeneity. Session Frequency emerged as a significant and robust predictor of antidepressant response. The model explained nearly all between-study variance [R^2=98.8%, QM (df=4)=59.08, p<0.001], with substantial reductions in residual heterogeneity (τ^2=0.31; I^2=57.8%). Trials administering two to five sessions showed large and statistically significant positive effects (*e.g.*, 2 sessions: β=6.25, p<0.001), indicating that a small number of high-quality sessions may be optimal. Intriguingly, a more intensive six-session protocol was associated with a significant attenuation of effect (β=-7.85, p=0.012), suggesting either therapeutic saturation or potential interference from over-structuring the intervention. This non-monotonic pattern implies that the therapeutic yield of session frequency is unlikely to be linear and may be contingent on the interplay between dose, integration, and patient readiness. In contrast, mean participant age did not significantly predict effect size variability [QM (df=1)=1.22, p=0.27], with a negligible coefficient (β=0.31, 95% CI: -0.24 to 0.87) and no reduction in residual variance (R^2=0%; τ^2=17.60; I^2=98.0%). These results indicate that age, within the adult samples studied, does not appear to moderate psilocybin efficacy. However, given the limited number of trials contributing to this analysis (k=7), a lack of power cannot be excluded. Similarly, follow-up duration showed no significant association with treatment effect (p=0.40), with the model explaining none of the observed heterogeneity (R^2=0%). The modest regression slope (β=0.10) and elevated τ^2 (30.18) suggest that changes in efficacy over time, whether enduring or transient, could not be meaningfully captured across the studied interval (1 to 60 weeks). These findings do not preclude delayed therapeutic trajectories but underscore the need for more consistent, long-term follow-up protocols to enable rigorous temporal modeling. Collectively, these results highlight session frequency as a clinically actionable moderator, while casting doubt on the predictive utility of age and follow-up interval in shaping treatment response. They also reinforce the necessity of designing future trials with precision around session dose and structure to optimize and stabilize therapeutic outcomes (*Supplementary Table 4*).

Sensitivity analyses and study influence diagnostics

To evaluate the structural stability of the meta-analytic findings, we performed a suite of sensitivity analyses focused on identifying whether specific trials disproportionately shaped the pooled effect. These included iterative leave-one-out procedures and comprehensive influence diagnostics encompassing standardized residuals, leverage, Cook's distance, and impact on model variance. The leave-one-out analysis indicated that the direction and magnitude of the pooled standardized mean difference remained largely invariant across exclusion permutations (*Supplementary Table 5*). No single study, when removed, shifted the overall result to statistical significance or altered its interpretive trajectory. This consistency suggests that the overall finding is not an artifact of undue weighting by any single trial and remains robust against individual study removal. Influence diagnostics plotted in Figure 4 revealed that two trials indexed here as Studies 6 and 7 exerted disproportionate influence on multiple model parameters. These studies produced the most extreme standardized residuals and exceeded critical thresholds on Cook's distance, hat values, and DFITS. Their exclusion markedly reduced between-study variance (τ^2.del) and improved model homogeneity (QE.del), pointing to their substantial contribution to overall heterogeneity. Importantly, these findings resonate with earlier observations of funnel plot asymmetry and underscore the necessity of interpreting these trials with heightened scrutiny. Nevertheless, excluding these studies did not yield a materially different pooled effect nor reduce heterogeneity to levels consistent with homogeneity. Their influence, while statistically salient, does not singularly distort the direction or non-significance of the aggregated outcome. Taken together, these diagnostics affirm the resilience of the main findings to analytic perturbation, while also highlighting trials that disproportionately contribute to variance. The results reinforce the importance of harmonized design standards, rigorous trial selection, and prespecified sensitivity modeling in future meta-analyses of psychedelic-assisted therapies.

OPEN ACCESS

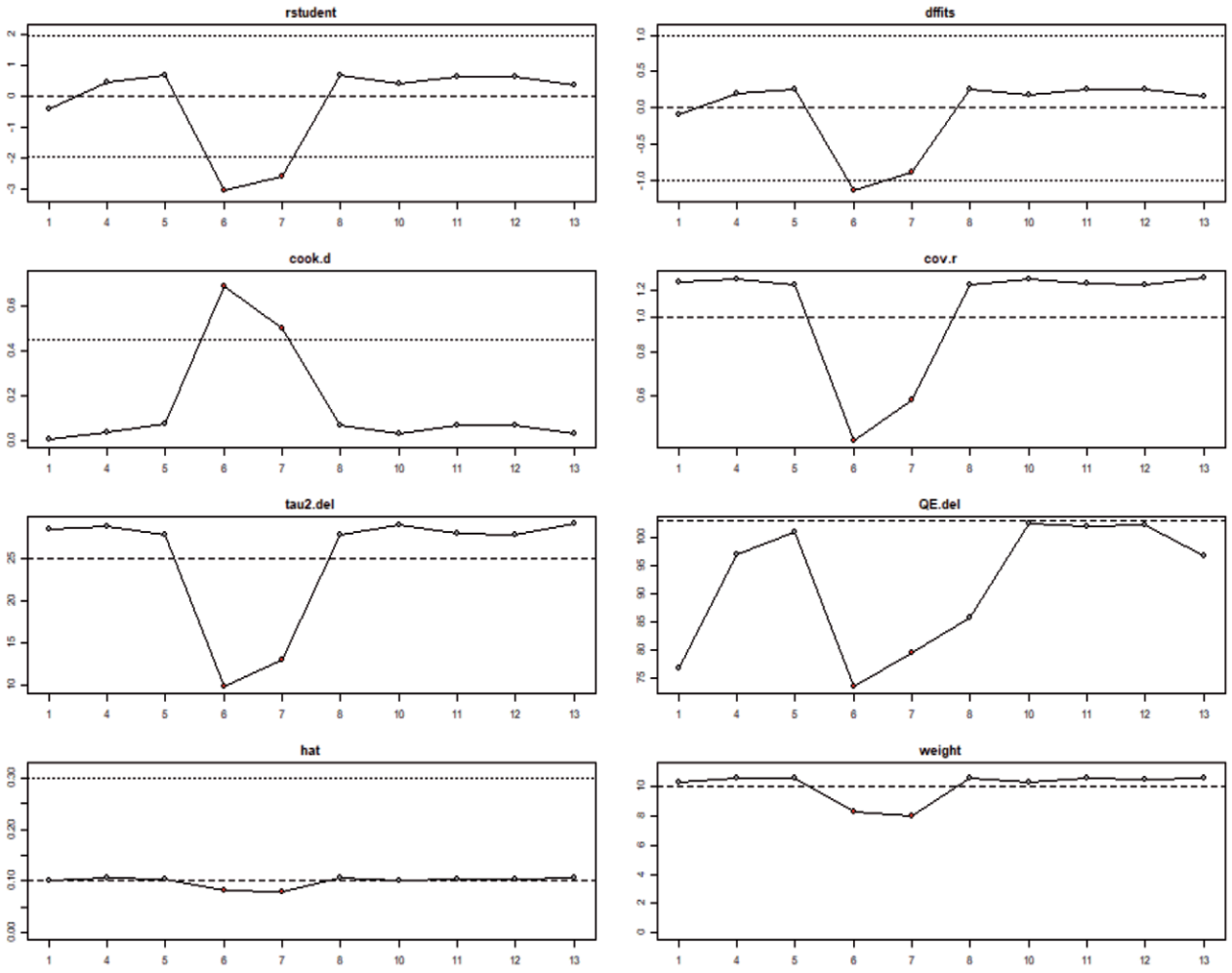

Figure 4. Influence diagnostics plot assessing leverage, standardized residuals, Cook's distance, and heterogeneity shifts across included studies.

Discussion

The present meta-analysis offers the most granular synthesis to date of psilocybin-assisted psychotherapy in clinically diagnosed major depressive and TRD populations. Contrary to early-phase enthusiasm, our findings do not support a uniform antidepressant effect across studies. Instead, the collective evidence reveals a heterogeneous and context-contingent pattern of therapeutic response, suggesting that psilocybin's clinical utility is not categorical but conditional. The pooled estimate did not reach statistical significance, with CIs spanning both clinically meaningful benefit and potential null effects. However, this ambiguity is not solely attributable to low statistical power. Rather, the observed heterogeneity, exceeding 95%, reflects genuine divergence in trial-level outcomes, not random sampling error. This pattern signals that the efficacy of psilocybin is unlikely to be generalizable across all populations and study designs, and must instead be interpreted as dependent on discrete clinical and methodological contexts. Crucially, the sources of this heterogeneity warrant closer examination. While we identified control group type and session frequency as key moderators, the potential collinearity or interaction between these factors—such as whether intensive protocols are more likely deployed alongside particular comparators, or whether lapses in blinding may confound apparent session effects—should be acknowledged. Future analyses would benefit from modeling these variables jointly to clarify their independent and interactive contributions to outcome variance.

Among the most consequential findings was the substantial moderating role of the control condition. Trials employing waitlist or inactive comparators consistently produced larger effect sizes than those using active pharmacologic controls, such as escitalopram. This asymmetry underscores the influence of nonspecific therapeutic factors—expectancy, therapeutic milieu, and symbolic framing of the psychedelic session—that are attenuated when psilocybin is benchmarked against an established antidepressant. The finding mirrors prior concerns in the psychedelic literature, where contextual amplification of perceived benefit complicates pharmacologic attribution.[12,15] Importantly, these results highlight the limitations of waitlist and low-dose controls in the context of psychedelic therapy, particularly in relation to blinding integrity and participant expectations. Expectancy effects and the unique therapeutic context may inflate observed efficacy with inactive controls. To minimize these confounds in future research, the adoption of active placebos with

OPEN ACCESS

psychoactive properties, enhanced blinding procedures, and systematic assessment of expectancy and therapeutic milieu is recommended. Additionally, outcome measures should be standardized to enable cross-study comparison and meta-analytic reliability.

Session frequency also emerged as a statistically significant and clinically meaningful moderator. Two-to-five session protocols were associated with the strongest effects, while more intensive six-session models paradoxically yielded weaker or even negative outcomes. This inverse association may reflect a non-linear therapeutic arc whereby experiential overload, diminished novelty, or insufficient integration may curtail psychological consolidation of earlier gains. Beyond "experiential overload", relevant psychotherapy process literature suggests that integration fatigue, participant burnout, or diminishing returns from repeated acute psychedelic experiences may also play a role. The optimal session number may vary by patient characteristics such as severity, chronicity, or comorbidity, and the observed non-linearity could reflect limitations of current trial designs rather than a universal principle. Further research is needed to delineate these factors and tailor session protocols to individual needs.[10,28]

Importantly, neither participant age nor follow-up duration significantly explained outcome variance. However, interpretive caution is warranted: the limited range of follow-up durations and small sample sizes in existing studies constrain our ability to draw robust conclusions about long-term durability or age-related differences. Larger, longer-term studies are needed to more definitively determine whether psilocybin's acute effects are truly age-independent or subject to attenuation over time. As such, our finding does not exclude the possibility of age or time effects, but highlights the field's current methodological limits.

Evidence of small-study effects, as revealed through Egger's test and funnel plot asymmetry, adds further caution to interpretation. Several influential trials exhibited outsized effects and contributed disproportionately to model heterogeneity. Their exclusion in sensitivity analysis did not reverse the overall null finding but did reduce variability, indicating that the field is particularly vulnerable to overestimation of effect magnitude in underpowered studies. To counter selective reporting and publication bias in future research, we strongly advocate for universal trial registration, full outcome reporting, and pre-specified analysis plans. These measures will increase transparency and reproducibility, and help ensure that effect estimates are not inflated by selective reporting or small-study effects.

Taken together, these results do not dismiss psilocybin's potential as an antidepressant, but they do delimit its conditions of efficacy. Psilocybin is not a panacea; its therapeutic impact appears highly sensitive to the intersection of pharmacology, psychotherapy, and trial architecture. Effective deployment of this intervention may therefore require not only a biologically active compound but also a highly controlled therapeutic scaffold capable of leveraging its acute experiential effects into sustained psychological transformation. Theoretically, these findings contribute to ongoing debates about the mechanisms of psychedelic action. The pronounced context-dependence observed in our synthesis may suggest that psychotherapeutic and expectancy-related mechanisms are at least as influential as direct neurobiological effects. This supports integrative models that emphasize both pharmacologic and contextual/psychotherapeutic contributions to clinical outcome. Further mechanistic studies, including patient-level meta-analyses and neurobiological substudies, are needed to clarify the interplay between drug, set, and setting.

Clinically, these findings argue for tighter standardization of session protocols, integration practices, and comparator conditions in future trials. Core components for future trial design should include careful selection of comparator arms (preferably active placebos or established antidepressants), standardized session structure and integration protocols, robust blinding procedures, and systematic assessment of expectancy effects. The field would benefit from multisite, adequately powered randomized controlled trials (RCTs), as well as mechanistic substudies and harmonized outcome measures to maximize generalizability and clarify sources of heterogeneity. From a regulatory standpoint, they justify cautious optimism while reinforcing the need for large-scale, multisite RCTs that can isolate true pharmacologic signal from contextual noise. Ultimately, the present synthesis reframes psilocybin not as a uniformly efficacious intervention, but as a mechanistically intriguing and clinically modifiable adjunct. Its utility in depression will likely depend not just on dose or diagnosis, but on the intentional design of the therapeutic container in which it is delivered.

Conclusions

This meta-analysis provides a nuanced and methodologically rigorous synthesis of clinical trials investigating psilocybin-assisted therapy for MDD and TRD. The evidence does not support a uniform antidepressant effect of psilocybin across all settings; rather, therapeutic outcomes are highly contingent on trial context, particularly the choice of comparator and the number of treatment sessions. Trials utilizing inactive or waitlist controls report larger effect sizes than those with active pharmacological comparators, highlighting the potential influence of expectancy and nonspecific therapeutic factors. Optimal efficacy appears to be associated with two-to-five session protocols, while more intensive regimens may yield diminishing or negative returns. Neither participant age nor follow-up duration significantly moderated outcomes, although limitations in study size and follow-up intervals restrict conclusions about long-term durability. Evidence of small-study effects and publication bias further tempers enthusiasm, suggesting that effect sizes in underpowered trials may be inflated. Ultimately, psilocybin should not be viewed as a universally effective antidepressant, but as a promising adjunct whose efficacy depends on careful integration of pharmacological, psychotherapeutic, and methodological elements. Future research should prioritize standardized protocols, robust control conditions, and adequately powered, multisite randomized trials to clarify the true therapeutic potential of psilocybin in depression.

References

1. World Health Organization. Depression. 2021. Available from: https://www.who.int/news-room/fact-sheets/detail/depression
2. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006;163:1905-17.
3. Carhart-Harris RL, Goodwin GM. The therapeutic potential of psychedelic drugs: past, present, and future. Neuropsychopharmacology 2017;42:2105-13.
4. Johnson MW, Garcia-Romeu A, Cosimano MP, Griffiths RR. Pilot study of the 5-HT2AR agonist psilocybin in the treatment of tobacco addiction. J Psychopharmacol 2014;28:983-92.
5. Griffiths RR, Johnson MW, Carducci MA, et al. Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: A randomized double-blind trial. J Psychopharmacol 2016;30:1181-97.
6. Ross S, Bossis A, Guss J, et al. Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: A randomized controlled trial. J Psychopharmacol 2016;30:1165-80.
7. Carhart-Harris RL, Erritzoe D, Williams T, et al. Neural corre-

OPEN ACCESS

lates of the psychedelic state as determined by fMRI studies with psilocybin. Proc Natl Acad Sci U S A 2012;109:2138-43.
8. Vollenweider FX, Preller KH. Psychedelic drugs: neurobiology and potential for treatment of psychiatric disorders. Nat Rev Neurosci 2020;21:611-24.
9. Davis AK, Barrett FS, May DG, et al. Effects of psilocybin-assisted therapy on major depressive disorder: a randomized clinical trial. JAMA Psychiatry 2020;78:1-9.
10. Gukasyan N, Davis AK, Barrett FS, et al. Efficacy and safety of psilocybin-assisted treatment for major depressive disorder: prospective 12-month follow-up. J Psychopharmacol 2022;36:151-8.
11. Carhart-Harris R, Giribaldi B, Watts R, et al. Trial of psilocybin versus escitalopram for depression. N Engl J Med 2021;384:1402-11.
12. Reiff CM, Richman EE, Nemeroff CB, et al. Psychedelics and psychedelic-assisted psychotherapy. Am J Psychiatry 2020;177:391-410.
13. Rucker JJH, Iliff J, Nutt DJ. Psychiatry and the psychedelic drugs: past, present, and future. Neuropharmacology 2018;142:200-18.
14. Johnson MW, Garcia-Romeu A, Griffiths RR. Long-term follow-up of psilocybin-facilitated smoking cessation. Am J Drug Alcohol Abuse 2017;43:55-60.
15. Studerus E, Gamma A, Kometer M, Vollenweider FX. Prediction of psilocybin response in healthy volunteers. PLoS One 2012;7:e30800.
16. Yao Y, Guo D, Lu TS, et al. Efficacy and safety of psychedelics for the treatment of mental disorders: a systematic review and meta-analysis. Psychiatry Res 2024;335:115886.
17. Asebey EJ, O'Reilly JT. FDA law and the psychedelic renaissance. FDLI Update. 2024. Available from: https://heinonline.org/HOL/LandingPage?handle=hein.journals/fdliup2024&div=8&id=&page=
18. Agrawal M, Richards W, Beaussant Y, et al. Psilocybin-assisted group therapy in patients with cancer diagnosed with a major depressive disorder. Cancer 2024;130:1137-46.
19. Goodwin GM, Croal M, Feifel D, et al. Psilocybin for treatment resistant depression in patients taking a concomitant SSRI medication. Neuropsychopharmacology 2023;48:1492-9.
20. von Rotz R, Schindowski EM, Jungwirth J, et al. Single-dose psilocybin-assisted therapy in major depressive disorder: a placebo-controlled, double-blind, randomised clinical trial. EClinicalMedicine 2022;56:101809.
21. Henry J, Giribaldi B, Nutt DJ, et al. The effects of psilocybin therapy versus escitalopram on cognitive bias: a secondary analysis of a randomized controlled trial. medRxiv 2025;2025-03.
22. Raison CL, Sanacora G, Woolley J, et al. Single-dose psilocybin treatment for major depressive disorder: a randomized clinical trial. JAMA 2023;330:843-53.
23. Back AL, Freeman-Young TK, Morgan L, et al. Psilocybin therapy for clinicians with symptoms of depression from frontline care during the COVID-19 pandemic: a randomized clinical trial. JAMA Netw Open 2024;7:e2449026.
24. Barba T, Buehler S, Kettner H, et al. Effects of psilocybin versus escitalopram on rumination and thought suppression in depression. BJPsych Open 2022;8:e163.
25. Erritzoe D, Barba T, Spriggs MJ, et al. Effects of discontinuation of serotonergic antidepressants prior to psilocybin therapy versus escitalopram for major depression. J Psychopharmacol 2024;38:458-70.
26. Rosenblat JD, Meshkat S, Doyle Z, et al. Psilocybin-assisted psychotherapy for treatment resistant depression: a randomized clinical trial evaluating repeated doses of psilocybin. Med 2024;5:190-200.
27. Goodwin GM, Aaronson ST, Alvarez O, et al. Single-dose psilocybin for a treatment-resistant episode of major depression: impact on patient-reported depression severity, anxiety, function, and quality of life. J Affect Disord 2023;327:120-7.
28. Haikazian S, Chen-Li DCJ, Johnson DE, et al. Psilocybin-assisted therapy for depression: a systematic review and meta-analysis. Psychiatry Res 2023;329:115531.

Online supplementary material:

Supplementary Table 1. Pooled standardized effect size, variance estimates, and heterogeneity diagnostics from the random-effects meta-analysis model.

Supplementary Table 2. Egger's regression test results for asymmetry in the funnel plot.

Supplementary Table 3. Mixed-effects meta-regression models evaluating the impact of blinding methodology and control condition on treatment effect heterogeneity.

Supplementary Table 4. Meta-regression results for continuous moderators of psilocybin efficacy.

Supplementary Table 5. Leave-one-out analysis and model robustness tests evaluating the impact of individual studies on the pooled effect estimate.

Received: 20 March 2026; Accepted: 23 March 2026.

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

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

Ethics approval and consent to participate: not applicable.

Informed consent: not applicable.

Patient consent for publication: not applicable.

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

Publisher's note: all claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).