nature mental health

Analysis

https://doi.org/10.1038/s44220-026-00630-8

A living systematic review, meta-analysis and open-data resource of randomized controlled trials of psilocybin treatment for symptoms of depression

Received: 21 August 2025

Accepted: 10 March 2026

Published online: 06 April 2026

Check for updates

S. Parker Singleton[1,2,12] ✉, Brooke L. Sevchik[1,2,12], Analiese Lahey[1,2], Pim Cuijpers[3,4], Mathias Harrer[3,5], Megan T. Jones[6], Sandeep M. Nayak[7], Eric C. Strain[7], Simon N. Vandekar[6], Robert H. Dworkin[8], J. Cobb Scott[9,10,13] & Theodore D. Satterthwaite[1,2,11,13] ✉

Depression is a major cause of disability worldwide, motivating interest in psilocybin as a potential treatment. Here we present a living systematic review and open-data meta-analytic resource on psilocybin treatment for depressive symptoms. In this initial release, 15 randomized controlled trials comprising 801 participants are included in the database, with 12 of those studies included in our primary model (n = 585) using inverse-variance random-effects modeling of standardized mean differences on primary outcomes. Compared with control conditions, psilocybin showed a greater reduction in depression scores (Hedges' g = −0.90). This systematic review and meta-analysis suggests that psilocybin-assisted therapy results in substantial decreases in depressive symptoms across studies to date. However, many studies have small sample sizes or risk of bias. This living systematic review, meta-analysis, database and online dashboard (spres.io) will continue to be updated as evidence emerges, providing a valuable resource for researchers in a rapidly evolving field.

Depression is a burden on health worldwide, with a growing prevalence and rising economic costs[1]. This high burden has motivated substantial public and private investment in the development of new treatments, including treatment with psychedelics[2]. Psilocybin, a serotonergic psychedelic, has seen renewed interest over the past decade driven by emerging evidence of its potential to treat conditions such as depression[3]. Whereas conventional pharmacotherapies for depression require daily dosing and may require weeks to take effect[4], psilocybin therapy typically centers around one or two dosing sessions, with antidepressant effects hypothesized to last long after the drug has been metabolized. To date, trials assessing psilocybin therapy administer the drug alongside psychotherapy or psychological support sessions in the days or weeks leading up to and following the dosing sessions, which occur under professional supervision and can last for 6–8 hours[5].

Over 130 clinical trials on the therapeutic potential of psilocybin have been initiated in the past two decades, sponsored by over 100 different institutions, and there is estimated to be a potential market size of $10.75 billion for the clinical use of psychedelics by 2027[6]. Thirty-nine of these initiated trials are for treatment-resistant depression or major depressive disorder, for which the US Food and Drug Administration has granted psilocybin Breakthrough Therapy designation to accelerate the approval process. The largest published trial to date assessed a single dose of psilocybin for a treatment-resistant episode of major depression in 233 individuals, finding that 25 mg, but not 10 mg, of psilocybin reduced depression scores significantly more than a 1 mg dose at 3 weeks[7].

A full list of affiliations appears at the end of the paper. ✉e-mail: parker.singleton@pennmedicine.upenn.edu; sattertt@pennmedicine.upenn.edu

Given the rapidly growing evidence base, systematic reviews on the topic[8–16] quickly become outdated, failing to capture the most recent findings. In response, we present a living systematic review on psilocybin treatment for depressive symptoms, ensuring that the latest findings are always reflected in the data synthesis. In addition, our continuously maintained database of effect sizes will be released using the Metapsy infrastructure (metapsy.org) and summarized in an interactive online dashboard (metapsy.org/sypres/psilocybin-depression). This dashboard allows scientists, clinicians and the broader public to retrieve the current evidence on the efficacy of psilocybin for depression. Moreover, it allows individuals to examine the influence of inclusion criteria, analysis choices and individually selected subgroups on the variability of study-level effects across this expanding literature. Herein, we describe our initial synthesis of these data. Our review will be updated at least annually; our data, code and results will be available on our public website: *SYPRES* (Synthesis of Psychedelic Research Studies; sypres.io)[17].

Results

After searching PubMed, Embase, PsycINFO, Web of Science, Scopus and the reference lists of systematic reviews retrieved from the searches, we identified 8,174 reports. Of these, 86 passed initial title and abstract screening and were reviewed as full-text articles. Twelve studies[7,18–28] met inclusion criteria for our primary model. An additional three studies[29–31] were included in the database and our sensitivity analyses on expanded inclusion criteria. Complete information about study identification and screening is included in Fig. 1. Our database consists of 220 effect sizes generated from 15 studies, covering multiple depression instruments, outcome types and time points (from baseline to over 6 months). This database has been publicly released and serves as the basis of our analyses that follow. All studies included a psychotherapy or psychological support component before, during and after psilocybin dosing (see 'Psychotherapy or psychological support' in the Supplementary Information). See Table 1 for detailed study characteristics.

Risk-of-bias ratings vary across studies

Overall, of the 15 studies in the database, 4 were determined to have a high risk of bias, 7 studies had some concerns, and 3 studies were deemed to have an overall low risk of bias according to Cochrane's Risk of Bias 2.0 tool (Table 2). Rosenblat et al.[24] received some concerns for randomization due to the significant imbalance in baseline depression severity between the two arms. This imbalance favors the comparator in our effect size calculation for this study. Krempien et al.[31] did not receive a risk-of-bias rating as the results have not been peer reviewed.

Psilocybin treatment significantly reduces depression symptoms compared with control conditions

To assess psilocybin's impact on continuous measures of depression symptoms, we performed inverse-variance random-effects modeling on standardized mean difference (Hedges' g) effect sizes calculated from continuous outcomes for each study. See Methods ('Effect size calculation' and 'Primary analysis on continuous outcomes') for details. The analysis on the 12 studies included in the primary model showed a statistically significant reduction in depression scores after psilocybin treatment compared with control conditions (Fig. 2; Hedges' $g = -0.90$ (-1.26; -0.55), $P < 0.001$, number of studies in the analysis (k) = 12, $n = 585$), with moderate between-study heterogeneity (estimated between-study variance (tau^2) = 0.12 (0.01; 1.01), percentage of variance due to between-study heterogeneity (I^2) = 53.9% (11.4%; 76.0%)). Visual inspection of a funnel plot (Supplementary Fig. 1) revealed limited asymmetry, and an Egger's test did not find small study effects (intercept = -1.58 (-3.88; -0.73), $t = -1.34$, $P = 0.21$).

Psilocybin's effects are rapid and consistent over several weeks

To assess psilocybin's effects independent of measurement time point, we applied a three-level correlated and hierarchical effects (CHE) meta-analysis model on 37 effect sizes ranging from 1 to 190 days post-dosing generated from the 12 studies included in our primary analysis. See Methods ('Three-level correlated and hierarchical effects meta-analysis') for more details. Our three-level CHE model revealed an overall significant decrease in depression scores with psilocybin compared with the control conditions (Hedges' $g = -0.82$ (-1.08; -0.57), $P < 0.001$, $k = 12$, $n = 619$, $\text{tau}^2 = 0.13$ (0.02; 0.32), $I^2 = 64.5\%$ (25.4%; 82.2%)), ensuring that our results were not sensitive to the time point studied. These results were consistent across a range of within-study correlation coefficients (Supplementary Fig. 2). Adding time since final dose as a continuous predictor to our model (Supplementary Fig. 3) revealed a significant effect favoring psilocybin immediately following dosing (intercept = -0.88 (-1.15; -0.61), $P < 0.0001$) that was stable over time (slope = 0.0019 (-0.0009; 0.0046), $P = 0.17$).

Higher response and remission rates after psilocybin treatment

To assess psilocybin's impact on dichotomous measures of response to treatment and remission, we performed inverse-variance random-effects modeling on risk ratios (RRs) calculated from raw event data. See Methods ('Effect size calculation' and 'Meta-analysis on dichotomous outcomes') for more details. In addition to differences in symptoms of depression, we found evidence for statistically significant greater treatment response with psilocybin compared with control conditions (Supplementary Fig. 4; RR = 2.77 (1.93; 3.97), $P = 0.001$, $k = 5$, $n = 380$), with low between-study heterogeneity ($\text{tau}^2 = 0.00$ (0.00; 0.96); $I^2 = 0.00\%$ (0.00%; 79.20%)). We also found that there were significantly higher remission rates with psilocybin compared with control conditions (Supplementary Fig. 5; RR = 4.04 (3.39; 4.82), $P < 0.001$, $k = 6$, $n = 405$), with low between-study heterogeneity ($\text{tau}^2 = 0.00$ (0.00; 0.00); $I^2 = 0.00\%$ (0.00%; 74.62%)).

Subgroup-specific analyses reveal open-label and crossover designs as high sources of heterogeneity

To assess potential sources of heterogeneity in the literature, we performed a series of subgroup-specific analyses. See Methods ('Subgroup-specific and sensitivity analyses') for more details. Hedges' g values of each subgroup-specific analysis were largely similar to the primary model, with four out of five analyses showing significant results (Fig. 3). Notably, subgroup-specific analyses excluding either open-label studies or crossover studies had substantially lower between-study heterogeneity ($\text{tau}^2 = 1.3 \times 10^{-6}$ (0; 0.34) and 1.5×10^{-6} (0; 1.10), respectively). However, analyzing crossover studies on their own resulted in substantially higher heterogeneity ($\text{tau}^2 = 0.44$ (0.05; 6.00)). It is worth noting, however, that the five studies included here vary in other important factors as well ($k = 2$ open-label waitlist control, $k = 3$ advanced cancer). The subgroup-specific analysis including only studies that used a diagnosis of major depressive disorder (MDD) as an inclusion criteria had a similar, albeit nonsignificant, effect size compared with the main model (Hedges' $g = -0.93$ (-1.87; 0.01), $P = 0.051$, $k = 5$, $\text{tau}^2 = 0.33$ (0.04; 6.01), $n = 358$). This lack of significance may be due to the decreased power and higher level of heterogeneity compared with the main model. A post hoc Bayesian implementation of this MDD model found a Hedges' g posterior distribution centered at -0.83 (-1.36; -0.33).

Sensitivity analyses show significant and comparable results to the primary model

We also performed a series of sensitivity analyses that largely supported our primary results.

First, our model using expanded inclusion criteria showed a significant and comparable effect size (Supplementary Fig. 6; Hedges' $g = -0.88$ (-1.21; -0.56), $P < 0.001$, $k = 15$, $\text{tau}^2 = 0.13$ (0.03; 0.92), $n = 686$). We also ran a model removing statistical outliers[20], resulting in a significant effect size comparable to the primary model (Hedges' $g = -0.81$ (-1.07; -0.55), $P < 0.001$, $k = 11$, $n = 561$) and reduced heterogeneity ($\text{tau}^2 = 0.015$ (0.00; 0.54)). Furthermore, we replicated

https://doi.org/10.1038/s44220-026-00630-8

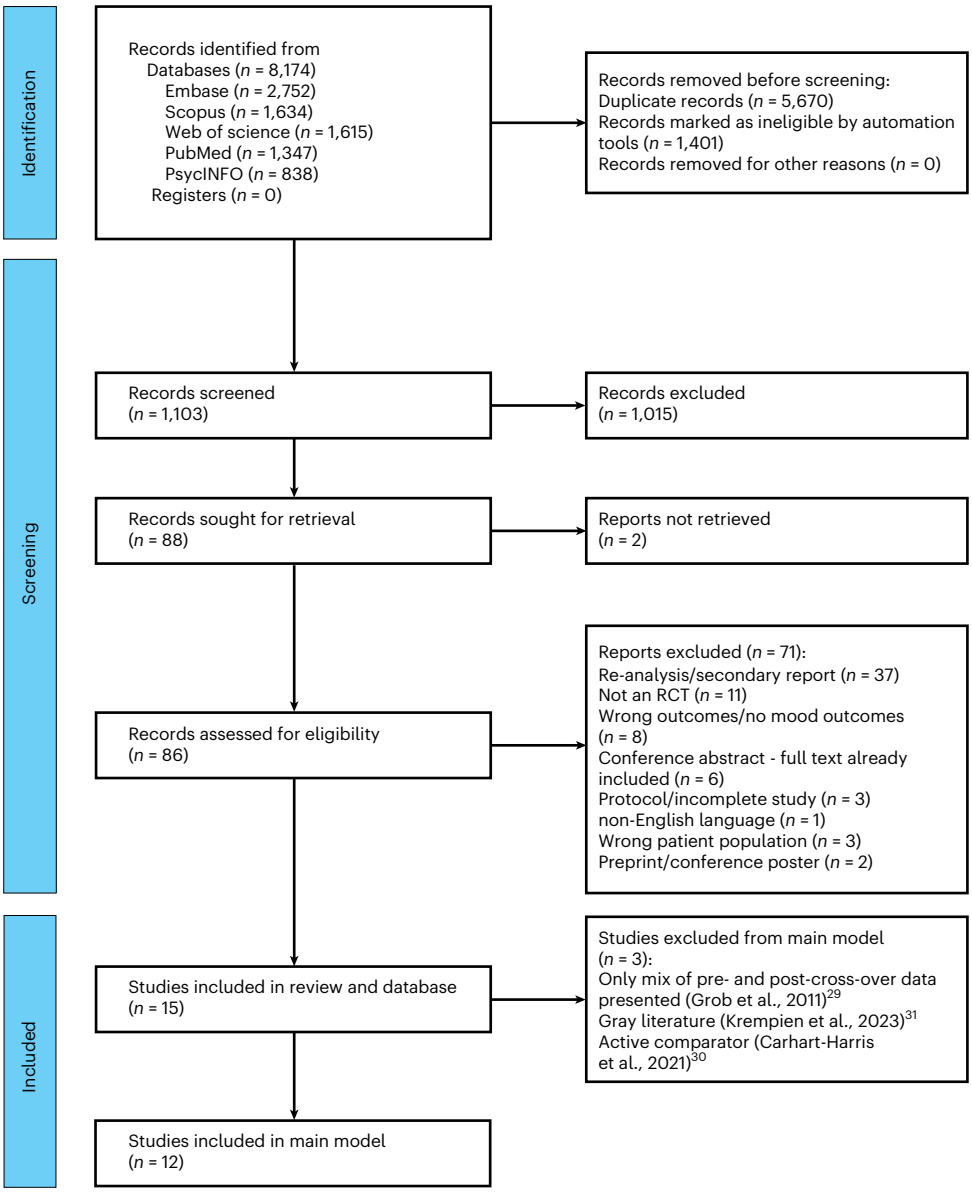

Fig. 1 | PRISMA 2020 flow diagram. 'Databases' includes 12 records that were merged with their parent studies from a total of 8,186 original records; see 'Search and selection' in the Supplementary Information for more details.

our primary analyses using a fixed-effect model on continuous outcomes (Hedges' $g = -0.84$ (−1.02; −0.67), $P < 0.001$, $k = 12$, $n = 585$), response outcomes (RR = 2.8 (2.05; 3.83), $P < 0.001$, $k = 5$, $n = 380$) and remission outcomes (RR = 4.05 (2.67; 6.15), $P < 0.001$, $k = 6$, $n = 405$). Separate models evaluating self-report (Hedges' $g = -1.00$ (−1.68; −0.31), $P = 0.010$, $k = 9$, $\text{tau}^2 = 0.49$ (0.15; 3.39), $n = 432$) and clinician-administered (Hedges' $g = -1.02$ (−1.60; −0.44), $P = 0.005$, $k = 7$, $\text{tau}^2 = 0.20$ (0.02; 2.35), $n = 438$) outcomes yielded comparable results, with the clinician-administered model having less heterogeneity despite fewer studies. Finally, our Bayesian analysis revealed a Hedges' g posterior distribution centered at −0.88 (−1.17; −0.60; Supplementary Fig. 7). All remaining sensitivity analyses showed significant and comparable results (Fig. 3).

Discussion

There is an urgent need for new therapies for depression[32,33]. Current evidence, synthesized here, suggests that psilocybin may have therapeutic potential for treating depression symptoms, with a large overall effect size. Compared with placebo, persons who received psilocybin had significantly lower depression scores and higher response and remission rates, with minimal variation in effect sizes by inclusion criteria or analysis choices, and these effects were rapid and consistent over several weeks. However, there was increased heterogeneity associated with study design and variability in risk of bias that should be considered when interpreting these findings. We present these results alongside an open-resource database, codebase and dashboard—providing a living resource that will be regularly updated as new evidence emerges.

Our work improves upon prior literature in several notable ways. First, while our results are consistent with previous meta-analyses, previous systematic reviews on the topic have quickly become outdated[8–16]. Second, in addition to our commitment to updating these results as a living review, our data and code are publicly accessible—maximizing transparency and reproducibility. Third, we also included numerous sensitivity analyses, which demonstrated results of similar magnitude across a variety of specific inclusion criteria and model parameters. Fourth, our subgroup-specific analyses identified potential sources of heterogeneity in the literature, including open-label studies and crossover study designs. Finally, our hierarchical effects meta-analysis

https://doi.org/10.1038/s44220-026-00630-8

Table 1 | Summary of RCTs on psilocybin for depressive symptoms

Study	*N*	Country	Age (s.d.)	Female (%)	Psych. use (%)	Patient diagnosis	Intervention	Control	Primary endpoint (weeks)
Primary meta-analysis									
Griffiths et al. (2016)[19]	56	USA	56.3 (10.5)	49	45	Advanced cancer	Single high dose psilocybin (22 or 30 mg 70 kg^{-1})	Single low dose psilocybin (1 or 3 mg 70 kg^{-1})	5
Ross et al. (2016)[18]	31	USA	56.3 (12.9)	62	55	Advanced cancer	Single dose psilocybin (0.3 mg kg^{-1})	Single dose niacin (250 mg)	7
Davis et al. (2021)[20]	27	USA	39.8 (12.2)	67	25	MDD	Two doses psilocybin (20 mg 70 kg^{-1}; 30 mg 70 kg^{-1})	Waitlist	4
Goodwin et al. (2022)[7]	158[a]	Multi	39.8 (12.2)	52	6	Treatment-resistant depression	Single high dose psilocybin (25 mg)	Single low dose psilocybin (1 mg)	3
Raison et al. (2023)[21]	104	USA	41.1 (11.3)	50	22	MDD	Single dose psilocybin (25 mg)	Single dose niacin (100 mg)	6
von Rotz et al. (2023)[22]	52	CH	36.8 (10.4)	63	31	MDD	Single dose psilocybin (0.215 mg kg^{-1})	Single dose placebo (mannitol)	2
Back et al. (2024)[23]	30	USA	38 (n/a)	50	7	COVID-19-related depression	Single dose psilocybin (25 mg)	Single dose niacin (100 mg)	4
Rosenblat et al. (2024)[24]	31	CA	44.4 (13.7)	39	19	Treatment-resistant depression or bipolar disorder II with current major depressive episode	Single dose psilocybin (25 mg)	Waitlist	2
Lewis et al. (2025)[25]	25	USA	43.6 (n/a)	72	48	COVID-19-related depression and burnout	Mindfulness-based stress reduction + single dose psilocybin (25 mg)	Mindfulness-based stress reduction + silent meditation retreat	2
Luquiens et al. (2025)[26]	30	FR	50.0 (n/a)	43	n/a	Alcohol use disorder	Two high doses psilocybin (25 mg)	Two low doses psilocybin (1 mg)	3
Rieser et al. (2025)[27]	40	CH	37.2 (11.9)	38	51	Alcohol use disorder	Single dose psilocybin (25 mg)	Single dose placebo (mannitol)	4
Ross et al. (2025)[28]	35	AUS	56.0 (13.6)	54	31.4	Advanced cancer	Single dose psilocybin (25 mg)	Single dose niacin (100 mg)	6
Total *N* (primary)	619	–	–		–	–	–	–	–
Mean (primary)	52	–	43.7	52	25	–	–	–	4.0
Expanded meta-analysis									
Grob et al. (2011)[29]	12	USA	36–58 (n/a)	92	67	Advanced cancer	Single dose psilocybin (0.2 mg kg^{-1})	Single dose niacin (250 mg)	2
Carhart-Harris et al. (2021)[30]	59	UK	41.2 (10.8)	34	27	MDD	Two doses psilocybin (25 mg) 3 weeks apart	Daily escitalopram (10–20 mg) for 6 weeks	3
Krempien et al. (2023)[31]	18[a]	USA	n/a	n/a	n/a	MDD	Single dose deuterated psilocybin derivative (16 mg)	Single dose placebo (unknown)	3
Total *N* (overall)	708	–	–	–	–	–	–	–	–
Mean (overall)	47	–	43.5	51	26	–	–	–	3.9

N is the number of participants randomized. Study endpoints are the reported primary endpoints for each study in weeks since the final dose. Mean summary statistics across studies were calculated as weighted averages. Overall mean age was calculated using the median age from Grob et al.[29]. USA, United States of America, UK, United Kingdom; CH, Switzerland; CA, Canada; Multi, multi-site; FR, France; AUS, Australia; n/a, data not available; MDD, major depressive disorder. The study by Goodwin et al.[7] was conducted at sites across the United States, Canada, United Kingdom and European Union. [a]*N* for these studies reflects the number of participants in the high- and low-dose groups. Goodwin et al.[7] and Krempien et al.[31] included 75 and 18 participants in the medium-dose arms, respectively, bringing the total *N* for the database to 801. These arms are included in the data but not used in our analyses, except the Goodwin et al.[7] medium-dose arm in the alternate dosing sensitivity analysis.

and meta-regression provide new results suggesting that psilocybin's effects over time are rapid and consistent over several weeks.

However, there remain substantial challenges in synthesizing evidence from currently available data[34]. Most psilocybin for depression trials to date have been small, with substantial variability in design and methods. In addition, study populations were often carefully selected and may not be representative of the general population that might receive treatment if psilocybin receives regulatory approval. As such, randomized controlled trials (RCTs) in the field often struggle with external validity[35–40]. Hence, future studies with larger treatment groups and expanded inclusion criteria will aid the generalizability of these findings.

https://doi.org/10.1038/s44220-026-00630-8

Table 2 | Additional study characteristics and risk-of-bias assessments using Cochrane's Risk of Bias 2.0 tool

Study	Design	Scale	Assessor	Risk-of-bias domains					
				Randomization	Deviations	Missingness	Measurement	Selection	Overall
Primary meta-analysis									
Griffiths et al. (2016)[19]	Double-blind placebo-controlled	GRID–HAMD	Non-independent study staff	Low	Low	Low	Some concerns	High	High
Ross et al. (2016)[18]	Double-blind placebo-controlled	BDI	Self-report	Low	Low	Low	Some concerns	Some concerns	Some concerns
Davis et al. (2021)[20]	Open-label waitlist-controlled	GRID–HAMD	Independent study staff	Low	Some concerns	Some concerns	Low	Low	Some concerns
Goodwin et al. (2022)[7]	Double-blind placebo-controlled	MADRS	Third party	Low	Low	Low	Low	Low	Low
Raison et al. (2023)[21]	Double-blind placebo-controlled	MADRS	Independent study staff	Low	Low	Low	Low	Low	Low
von Rotz et al. (2023)[22]	Double-blind placebo-controlled	MADRS	Non-independent study staff	Low	Low	Low	Some concerns	Low	Some concerns
Back et al. (2024)[23]	Double-blind placebo-controlled	MADRS	Independent study staff	Low	Low	Low	Low	Low	Low
Rosenblat et al. (2024)[24]	Open-label waitlist-controlled	MADRS	Non-independent study staff	Some concerns	Low	Low	High	Low	High
Lewis et al. (2025)[25]	Open-label therapy-controlled	QIDS–SR	Self-report	Low	Low	Low	Some concerns	Low	Some concerns
Luquiens et al. (2025)[26]	Double-blind placebo-controlled	BDI–II	Self-report	Low	High	Low	Some concerns	Low	High
Rieser et al. (2025)[27]	Double-blind placebo-controlled	BDI–II	Self-report	Low	Low	Some concerns	Some concerns	Low	Some concerns
Ross et al. (2025)[28]	Double-blind placebo-controlled	HADS–D	Self-report	Low	Low	Low	Some concerns	Low	Some concerns
Expanded meta-analysis									
Grob et al. (2011)[29]	Double-blind placebo-controlled	BDI	Self-report	Low	Some concerns	Low	Some concerns	High	High
Carhart-Harris et al. (2021)[30]	Double-blind placebo-controlled	QIDS–SR	Self-report	Low	Low	Low	Some concerns	Low	Some concerns
Krempien et al. (2023)[31]	Double-blind placebo-controlled	MADRS	Unknown	NA	NA	NA	NA	NA	NA

Randomization, bias due to randomization process; Deviations, bias due to deviations from intended interventions; Missingness, bias due to missing outcome data; Measurement, bias due to measurement of the outcome; Selection, bias due to selection of the reported results; Overall, overall risk of bias; BDI, Beck Depression Inventory; GRID–HAMD, Grid Hamilton Rating Scale for Depression; MADRS, Montgomery–Åsberg Depression Rating Scale; QIDS–SR, Quick Inventory of Depressive Symptomatology–Self-Report; HADS–D, Hospital Anxiety and Depression Scale–Depression.

Issues such as functional unblinding and expectancy effects also complicate interpretation. Double-blinding in psychedelic-assisted psychotherapy has been the topic of extensive discourse in the field[41–44]. The psychoactive effects of psilocybin make functional unblinding likely, and studies that have tested functional unblinding find that almost all participants and/or blinded study staff correctly guess group assignments[18,19,23,27]. In light of this, a recent meta-analysis comparing pre–post effect sizes of psilocybin treatment with open-label selective serotonin-reuptake inhibitor trials found the two therapies are nearly identical in their efficacy for depression[45]. In addition to increased expectancy in arms receiving psilocybin treatment, there may be a substantial underperformance in the placebo arm, as evidenced by an underperformance of placebo arms in psychedelic trials compared with placebo arms in trials of escitalopram (a selective serotonin-reuptake inhibitor)[44–47]. Given this potential for inflated effect sizes in psilocybin trials, our large pooled effect size should be interpreted with caution. In addition, we found either some concerns or high risk of bias in most studies, even without fully accounting for potential bias from functional unblinding or expectancy effects.

There are several limitations to this meta-analysis. First, the power of our results is limited by the small number of studies that met eligibility criteria. Second, available endpoint data from longer durations is sparse, and longer-duration studies are needed to determine the durability of psilocybin's effects beyond a few months. Further, although the linear meta-regression model applied for this analysis favors interpretability, this could obscure nonlinear relationships between depression symptoms and time after a psilocybin dose. Third, we were unable to examine the number of psilocybin doses as an explanatory variable given limited studies. For example, although Davis et al.[20] was identified as a statistical outlier, we are unable to conclude whether this is due to its administration of two psilocybin doses, its low number of participants or its waitlist design. As the variation in study design grows with the field, we plan to expand our exploration of these important variables with additional subgroup-specific and regression analyses. Fourth, while our subgroup-specific and sensitivity analyses emphasize the robustness of our findings, conducting multiple analyses increases the likelihood of false positive results. Finally, the role of psychotherapy or psychological support during psilocybin treatment sessions is a topic of substantial importance in the field[48–55]. All included studies in our review contained comparable amounts of patient contact before, during and after psilocybin dosing sessions. As such, we were unable to disambiguate the contributions that conjunctive psychotherapy and psychological support make in psilocybin's effects on depression symptoms. It is worth noting that improved reporting on

https://doi.org/10.1038/s44220-026-00630-8

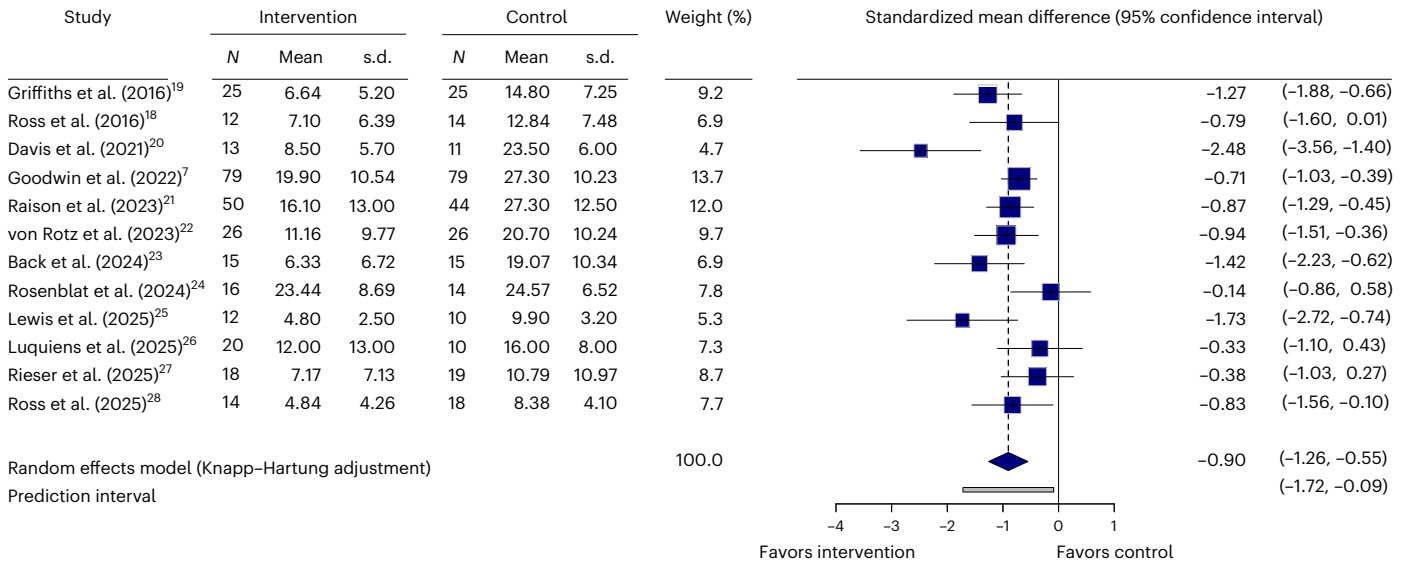

Fig. 2 | Primary meta-analysis on continuous outcome variables. Boxes represent Hedges' *g* for each study, and the lines extending from the box represent the 95% confidence interval around each effect size; the size of each box is proportional to its weight. The diamond at the bottom represents the pooled effect size (meta-analytic mean). The gray line at the bottom represents the prediction interval of the expected range of true effects in a new study. Heterogeneity, $I^2 = 53.9\%$, $\text{tau}^2 = 0.1174$, $P = 0.0134$.

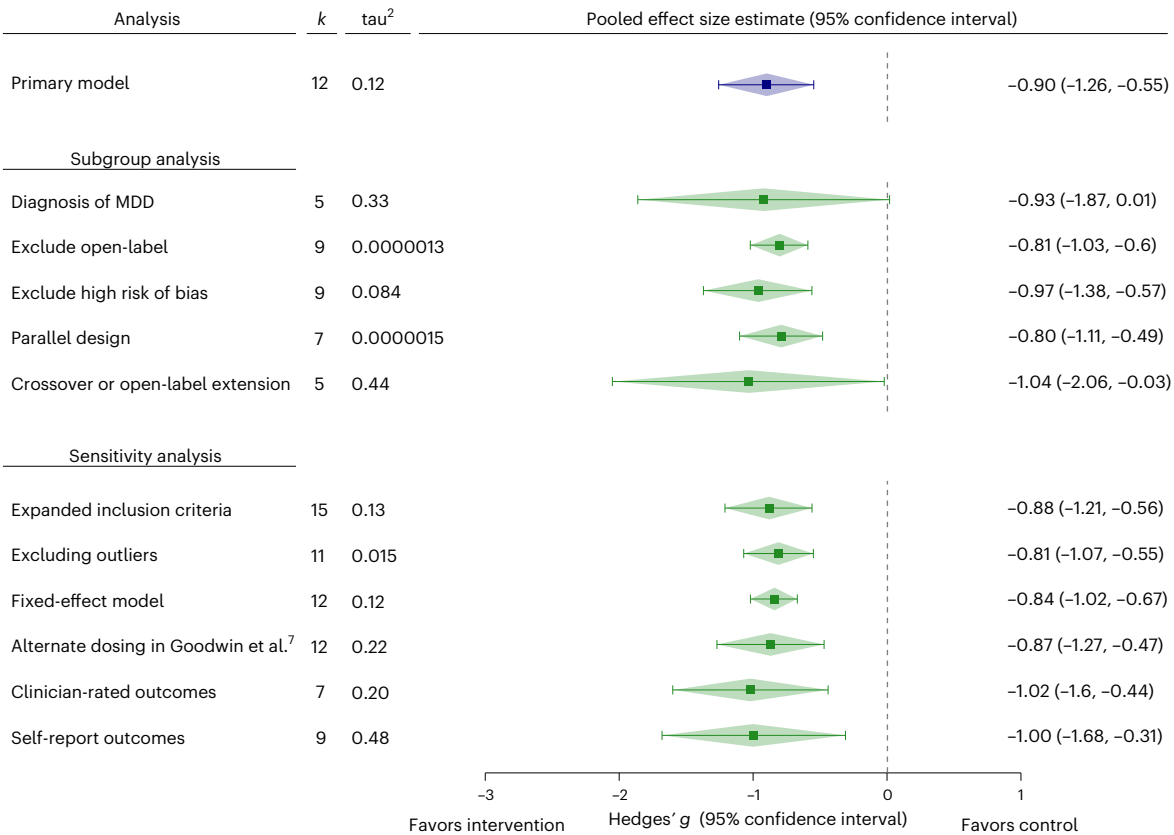

Fig. 3 | Pooled effect sizes for subgroup-specific and sensitivity analyses. Box and whiskers represent the meta-analytic mean and corresponding 95% confidence intervals for each subgroup-specific analysis. The pooled effect size from the primary model is presented at the top for comparison purposes.

the psychological interventions accompanying psychedelic therapy will improve replicability, generalizability and interpretations of this research[5]. As more studies become available, future supplemental analyses may explore this factor.

Within the context of these important limitations, our results suggest that psilocybin may be a promising treatment for symptoms of depression, although larger studies with expanded patient populations and rigorous methods are needed. As more RCTs are published, we will regularly update our SYPRES website and dashboard in a reproducible and transparent manner. As part of our SYPRES initiative, we will also conduct a series of future meta-analyses on other psychedelic therapies, including MDMA (3,4-methylenedioxymethamphetamine) for

post-traumatic stress disorder and psilocybin for anxiety[17]. This living systematic review and open science resource will provide a valuable and transparent resource for researchers, clinicians, policymakers and the public.

Methods

This meta-analysis is a pre-registered study under the SYPRES project, and the study protocol is available on PROSPERO (https://www.crd.york.ac.uk/PROSPERO/view/CRD42024584938). Please refer to the Supplementary Methods for detailed methodological information.

Eligibility criteria

We included randomized controlled clinical trials published in English in peer-reviewed journals comparing psilocybin for depressive symptoms with a comparator in adult (>18 years old) populations. To meet this criteria, the study population needed to include at least some individuals with elevated depressive symptoms; studies considering only healthy participants were not considered. Eligible interventions included any dose and formulation (natural or synthetic) of psilocybin or other prodrugs of psilocin intended to produce an alteration of subjective experience in the patient, with or without the conjunctive use of therapy (for example, microdosing studies were not included). Eligible comparators included any form of placebo with or without the conjunctive use of therapy, including low doses of the intervention drug (sub-threshold for experience alteration) and any dose of other psychotropics intended to improve blinding (without known therapeutic efficacy for depression). Eligible comparators also included control for spontaneous improvement via waitlist or usual care. In addition, due to the potential carry-over effects of a psilocybin dose, data from pre-crossover time points was required. To make our database more comprehensive, we also conducted sensitivity analyses where we expanded these criteria to include reports found in gray literature ($k = 1$; Krempien et al.[31]), studies presenting only a mixture of pre- and post-crossover data ($k = 1$; Grob et al.[29]) and studies comparing psilocybin with comparators with known therapeutic efficacy ($k = 1$; Carhart-Harris et al.[30]). However, these studies were not included in the primary analyses.

Study search and selection

We searched PubMed, Embase, PsycINFO, Web of Science, Scopus and the reference lists of systematic reviews retrieved from the searches. Expert research librarians assisted in crafting study search criteria. The final search was conducted on 31 October 2025 (see 'Search terms' in the Supplementary Information). Study screening and data extraction were independently performed by two reviewers. We assessed risk of bias using Cochrane's Risk of Bias 2.0 tool. We will update our search and the accompanying analyses and online resources at least annually for the next 5 years. This study's senior authors will be responsible for ongoing supervision. An updated publication may be prepared at a future date if highly relevant trials are published or the evidence base substantially shifts or matures.

Effect size calculation

We used each study's endpoint sample sizes, means and standard deviations for treatment and control to calculate Hedges' g for each study as the primary effect size measure for continuous outcomes. Our reported effect sizes therefore may differ from study reports using change-from-baseline values (see 'Study harmonization' in the Supplementary Information for full discussion). Hedges' g is a standardized mean difference (SMD). SMD is used as a summary statistic in meta-analysis when different measurements are used across studies to assess the same outcome (for example, different depression scales in this case). The SMD standardizes results from different measures to a uniform scale by expressing the size of the intervention effect relative to between-participant variability under the assumption that differences in standard deviations across studies are due only to differences in the measurement scales and not to real differences in variability among the populations studied[56]. For dichotomous outcomes, we calculated RRs from raw event data. RRs quantify how much an intervention multiplies the likelihood of an outcome occurring. For example, when an intervention has an RR of 3, this indicates that the outcome is 3 times more probable in the intervention arm than in the control arm. Likewise, an RR of 0.25 means that the probability of an outcome occurring in the intervention arm is one-fourth that in the control arm. An RR of 1 indicates equal probabilities in both arms. Therefore RRs provide an easily interpretable summary statistic of dichotomous outcomes[56]. All analyses except the three-level model (see 'Three-level correlated and hierarchical effects meta-analysis') employed one effect size per study. See 'Selection of effect sizes' in the Supplementary Information for details.

Meta-analyses

Primary analysis on continuous outcomes. We performed inverse-variance random-effects modeling of SMDs on primary outcomes (see 'Selection of effect sizes' in Supplementary Methods and Supplementary Tables 2 and 3 for selection of outcomes and time points for effect sizes). Between-study heterogeneity (tau^2) was calculated using the restricted maximum likelihood (REML) estimator[57] and the Q-profile method to calculate the confidence interval[58]. We employed the Knapp–Hartung adjustment to calculate the confidence interval for the pooled effect size[59]. This adjustment creates a more conservative confidence interval and P value that varies less with changes in heterogeneity variance. Funnel plots and Egger's test were used to assess small study bias, although the small number of studies limits the precision of these approaches[60].

Three-level correlated and hierarchical effects meta-analysis. To assess psilocybin's effects independent of measurement time point, we applied a three-level meta-analysis model on 37 effect sizes ranging from 1 to 190 days post-dosing generated from the 12 studies included in our primary analysis (see Supplementary Table 4 for full list of effect sizes). These effect sizes were limited to assessments on the primary instrument occurring at least 1 day after dosing and before any crossover between groups. Variance–covariance matrices of each study with two or more effect sizes were estimated using a within-study correlation coefficient (ρ) of 0.6, creating a CHE model, which is typically a good approximation for datasets with unknown and/or complex dependence structures[61]. To test the sensitivity of our results against this approximation, we recalculated Hedges' g as a function of ρ from 0 to 1 in 0.1 increments. Cluster-robust variance estimation was used to guard against potential model misspecification. Heterogeneity was calculated using the REML estimator with parametric bootstrapping (5,000 iterations) used to generate confidence intervals. To examine the consistency of effects over time, we additionally performed a linear meta-regression using the three-level CHE model by adding time since the final dose (in days) as a continuous predictor.

Meta-analysis on dichotomous outcomes. In addition to continuous outcomes, we evaluated dichotomous response and remission outcomes reported by five studies[7,18,19,21,22]. Response typically refers to the minimum change in a scale to signify clinically relevant improvement in symptoms, while remission typically refers to an endpoint measurement that falls below a specified cut-off for a diagnosis of depression. Both of these variables were defined and reported by each individual study. We used inverse-variance random-effects modeling of risk ratios. Between-study heterogeneity was calculated with the Paule–Mandel estimator[62], which is an alternative to REML with a good performance in analyzing dichotomous outcomes. Confidence intervals were again calculated using the Q-profile method for heterogeneity estimates and the Knapp–Hartung adjustment for pooled effect sizes.

Subgroup-specific and sensitivity analyses. To assess the robustness of our primary findings and explore potential sources of heterogeneity, we conducted a series of five subgroup-specific analyses for continuous outcomes. These analyses allowed us to evaluate the consistency of treatment effects across different study characteristics and methodological approaches:

1. Diagnosis of MDD: to assess whether participant diagnosis impacted treatment effects, we conducted a subgroup analysis limited to studies focused on patients with a formal diagnosis of MDD (includes Davis et al.[20], Goodwin et al.[7], Raison et al.[21], von Rotz et al.[22] and Rosenblat et al.[24]).
2. Exclude open-label: we evaluated the impact of excluding open-label studies (excluding Davis et al.[20], Rosenblat et al.[24] and Lewis et al.[25]).
3. Exclude high risk of bias: to assess the impact of study quality on outcomes, we excluded studies rated as having high risk of bias (excluding Griffiths et al.[19], Rosenblat et al.[24] and Luquiens et al.[26]).
4. Parallel design: we conducted a subgroup-specific analysis that included only parallel group studies (including Goodwin et al.[7], Raison et al.[21], von Rotz et al.[22], Back et al.[23], Lewis et al.[25], Luquiens et al.[26] and Rieser et al.[27]).
5. Crossover or open-label extension design: we conducted an analysis that included only crossover design studies and studies with open-label extension periods (including Griffiths et al.[19], Ross et al.[18], Davis et al.[20], Rosenblat et al.[24] and Ross et al.[28]).

The following sensitivity analyses were also performed:

1. Expanded inclusion criteria: we conducted an analysis with expanded eligibility criteria that incorporated all studies from the primary model plus three studies that were excluded from primary analyses (Eligibility criteria).
2. Excluding outliers: we repeated our primary meta-analysis on continuous outcomes after removing statistical outlier studies (for example, whose effect size confidence intervals do not overlap with the confidence interval of the pooled effect; Davis et al.[20]).
3. Fixed-effect model: we ran fixed-effect models as sensitivity analyses to compare with random-effects models. Fixed-effect models assume that the between-study variance (tau^2) is 0, such that all studies share a common true effect size. For our continuous model, we used a standard inverse-variance weighting fixed-effect model on standardized mean differences (Hedges' *g*). For our dichotomous model, we used a standard inverse-variance weighting fixed-effect model on the log risk ratio.
4. Alternate dosing in Goodwin et al.[7]: given that Goodwin et al.[7] employed a three-arm design comparing psilocybin at 25 mg and 10 mg doses against a 1 mg control, we conducted a subgroup-specific analysis substituting the 10 mg intervention arm for the 25 mg arm used in our primary analysis.
5. Clinician-rated outcomes: to examine whether the method of assessment influenced observed effects, we conducted a sensitivity analysis that included only clinician-administered depression assessments. MADRS was chosen as the preferred clinician-administered instrument for this analysis, followed by the GRID-HAM-D. Studies included were Goodwin et al.[7], Raison et al.[21], von Rotz et al.[22], Back et al.[23] and Rosenblat et al.[24] using the MADRS and Griffiths et al.[19] and Davis et al.[20] using the GRID-HAM-D.
6. Self-report outcomes: similarly, we conducted a sensitivity analysis that included only studies reporting self-report depression measures, regardless of whether these were the primary outcome measures. BDI was the preferred self-report instrument that was reported in all self-report studies except two for which only the QIDS–SR was available. Studies included were Ross et al.[18], Griffiths et al.[19], Davis et al.[20], Goodwin et al.[7], von Rotz et al.[22], Lewis et al.[25], Luquiens et al.[26], Rieser et al.[27] and Ross et al.[28].
7. Bayesian meta-analysis: we replicated our primary meta-analysis on continuous outcomes using a Bayesian implementation. We used 'weakly informative' prior distributions for both the main effect and the heterogeneity parameter *tau* that have been recommended by previous work[63,64]. The main effect prior was a normal distribution centered around 0, with a standard deviation of 1, while the *tau* prior was a half-normal distribution with a standard deviation of 0.5. This same approach was used in the post hoc Bayesian analysis of the MDD subgroup.

Software. Literature screening and data extraction were performed using Covidence. Meta-analyses were conducted using R (4.4.1) in RStudio (2024.04.2+764) with metapsyTools (1.0.12)[65], a package of helper functions for Metapsy that uses meta (7.0.0)[66], metafor (4.6.0)[67] and dmetar (0.1.0)[68] functions. The Bayesian meta-analysis was implemented using the bayesmeta (3.5)[69] R package.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The living database used for this analysis can be accessed via Metapsy (docs.metapsy.org/databases/depression-psiloctr/) or via Zenodo (https://doi.org/10.5281/zenodo.15714852)[70].

Code availability

Code for all analyses can be accessed on our website (sypres.io), or on our GitHub repository (https://github.com/PennLINC/sypres-docs) (release 26.0.0).

References

1. Proudman, D., Greenberg, P. & Nellesen, D. The growing burden of major depressive disorders (MDD): implications for researchers and policy makers. *PharmacoEconomics* **39**, 619–625 (2021).
2. Siegel, J. S., Daily, J. E., Perry, D. A. & Nicol, G. E. Psychedelic drug legislative reform and legalization in the US. *JAMA Psychiatry* **80**, 77–83 (2023).
3. Nutt, D., Erritzoe, D. & Carhart-Harris, R. Psychedelic psychiatry's brave new world. *Cell* **181**, 24–28 (2020).
4. Quitkin, F. M., Rabkin, J. G., Ross, D. & McGrath, P. J. Duration of antidepressant drug treatment: what is an adequate trial? *Arch. Gen. Psychiatry* **41**, 238–245 (1984).
5. Seybert, C. et al. Quality of reporting on psychological interventions in psychedelic treatments: a systematic review. *Lancet Psychiatry* **12**, 54–66 (2025).
6. Norring, S. & Spigarelli, M. The promise of therapeutic psilocybin: an evaluation of the 134 clinical trials, 54 potential indications, and 0 marketing approvals on ClinicalTrials.gov. *Drug Des. Devel. Ther.* **18**, 1143–1151 (2024).
7. Goodwin, G. M. et al. Single-dose psilocybin for a treatment-resistant episode of major depression. *N. Engl. J. Med.* **387**, 1637–1648 (2022).
8. Ko, K., Kopra, E. I., Cleare, A. J. & Rucker, J. J. Psychedelic therapy for depressive symptoms: a systematic review and meta-analysis. *J. Affect. Disord.* **322**, 194–204 (2023).
9. Haikazian, S. et al. Psilocybin-assisted therapy for depression: a systematic review and meta-analysis. *Psychiatry Res.* **329**, 115531 (2023).
10. Aghajanian, S. et al. Psilocybin for major depressive disorder: an updated systematic review and meta-analysis of randomized clinical trials. *J. Psychopharmacol* https://doi.org/10.1177/02698811241287542 (2024).

11. Fang, S., Yang, X. & Zhang, W. Efficacy and acceptability of psilocybin for primary or secondary depression: a systematic review and meta-analysis of randomized controlled trials. *Front. Psychiatry* **15**, 1359088 (2024).
12. Dawood Hristova, J. J. & Pérez-Jover, V. Psychotherapy with psilocybin for depression: systematic review. *Behav. Sci.* **13**, 297 (2023).
13. Li, L. J. et al. Psilocybin for major depressive disorder: a systematic review of randomized controlled studies. *Front. Psychiatry.* **15**, 1416420 (2024).
14. Metaxa, A. M. & Clarke, M. Efficacy of psilocybin for treating symptoms of depression: systematic review and meta-analysis. *Brit. Med. J.* **385**, e078084 (2024).
15. Watford, T. & Masood, N. Psilocybin, an effective treatment for major depressive disorder in adults—a systematic review. *Clin. Psychopharmacol. Neurosci.* **22**, 2–12 (2024).
16. Menon, V., Ramamurthy, P., Venu, S. & Andrade, C. Randomized controlled trials of psilocybin-assisted therapy in the treatment of major depressive disorder: systematic review and meta-analysis. *Acta Psychiatr. Scand.* https://doi.org/10.1111/acps.13778 (2024).
17. Singleton, S. P. et al. An initiative for living evidence synthesis in clinical psychedelic research. *Nat. Ment. Health* **3**, 3–5 (2025).
18. Ross, S. 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.* **30**, 1165–1180 (2016).
19. Griffiths, R. R. 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.* **30**, 1181–1197 (2016).
20. Davis, A. K. et al. Effects of psilocybin-assisted therapy on major depressive disorder: a randomized clinical trial. *JAMA Psychiatry* **78**, 481–489 (2021).
21. Raison, C. L. et al. Single-dose psilocybin treatment for major depressive disorder: a randomized clinical trial. *JAMA* **330**, 843–853 (2023).
22. von Rotz, R. et al. Single-dose psilocybin-assisted therapy in major depressive disorder: a placebo-controlled, double-blind, randomised clinical trial. *eClinicalMedicine* **56**, 101809 (2023).
23. Back, A. 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* **7**, e2449026 (2024).
24. Rosenblat, J. D. et al. Psilocybin-assisted psychotherapy for treatment resistant depression: a randomized clinical trial evaluating repeated doses of psilocybin. *Med* **5**, 190–200.e5 (2024).
25. Lewis, B. R. et al. Psilocybin-assisted group psychotherapy and mindfulness-based stress reduction for frontline healthcare provider COVID-19-related depression and burnout: a randomized controlled trial. *PLoS Med.* **22**, e1004519 (2025).
26. Luquiens, A. et al. Psilocybin in alcohol use disorder and comorbid depressive symptoms: results from a feasibility randomized clinical trial. *Addiction* https://doi.org/10.1111/add.70152 (2025).
27. Rieser, N. M. et al. Psilocybin-assisted therapy for relapse prevention in alcohol use disorder: a phase 2 randomized clinical trial. *eClinicalMedicine* **82**, 103149 (2025).
28. Ross, M. L. et al. Psilocybin-assisted psychotherapy for depression and anxiety associated with life threatening illness: a phase 2b randomized controlled trial. *Gen. Hosp. Psychiatry* **96**, 322–331 (2025).
29. Grob, C. S. et al. Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. *Arch. Gen. Psychiatry* **68**, 71–78 (2011).
30. Carhart-Harris, R. et al. Trial of psilocybin versus escitalopram for depression. *N. Engl. J. Med.* **384**, 1402–1411 (2021).
31. Krempien, S. et al. P226. Randomized, double-blind, placebo-controlled study to evaluate the safety and efficacy of CYB003, a deuterated psilocybin analog in patients with major depressive disorder. *Neuropsychopharmacology* **48**, 63–210 (2023).
32. Popa-Velea, O., Gheorghe, I., Trușescu, C. & Purcărea, V. Current challenges and pitfalls in the pharmacological treatment of depression. *J. Med. Life* **8**, 181–186 (2015).
33. Zhdanava, M. et al. The prevalence and national burden of treatment-resistant depression and major depressive disorder in the United States. *J. Clin. Psychiatry* **82**, 20m13699 (2021).
34. van Elk, M. & Fried, E. I. History repeating: guidelines to address common problems in psychedelic science. *Ther. Adv. Psychopharmacol.* **13**, 20451253231198466 (2023).
35. Lieberman, J. A. et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N. Engl. J. Med.* **353**, 1209–1223 (2005).
36. Stroup, T. S. et al. Effectiveness of olanzapine, quetiapine, risperidone, and ziprasidone in patients with chronic schizophrenia following discontinuation of a previous atypical antipsychotic. *Am. J. Psychiatry* **163**, 611–622 (2006).
37. Trivedi, M. H. et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am. J. Psychiatry* **163**, 28–40 (2006).
38. Warden, D., Rush, A. J., Trivedi, M. H., Fava, M. & Wisniewski, S. R. The STAR*D Project results: a comprehensive review of findings. *Curr. Psychiatry Rep.* **9**, 449–459 (2007).
39. Miklowitz, D. J. et al. Psychosocial treatments for bipolar depression: a 1-year randomized trial from the Systematic Treatment Enhancement Program. *Arch. Gen. Psychiatry* **64**, 419–426 (2007).
40. Parikh, S. V., LeBlanc, S. R. & Ovanessian, M. M. Advancing bipolar disorder: key lessons from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *Can. J.Psychiatry* **55**, 136–143 (2010).
41. Szigeti, B. & Heifets, B. D. Expectancy effects in psychedelic trials. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **9**, 512–521 (2024).
42. Schenberg, E. E. Who is blind in psychedelic research? Letter to the editor regarding: blinding and expectancy confounds in psychedelic randomized controlled trials. *Expert Rev. Clin. Pharmacol.* **14**, 1317–1319 (2021).
43. Muthukumaraswamy, S. et al. Reimagining psychedelic trials: theorising a pragmatic approach to complex intervention development. *OSF* https://doi.org/10.31234/osf.io/caup9_v1 (2025).
44. Butler, M., Jelen, L. & Rucker, J. Expectancy in placebo-controlled trials of psychedelics: if so, so what? *Psychopharmacology (Berl)* **239**, 3047–3055 (2022).
45. Williams, Z. J., Barnett, H. & Szigeti, B. Psychedelic therapy vs. antidepressants for the treatment of depression under equal unblinding conditions: a systematic review and meta-analysis. *JAMA Psychiatry* https://doi.org/10.1001/jamapsychiatry.2025.4809 (2026).
46. Hsu, T. W. et al. Comparative oral monotherapy of psilocybin, lysergic acid diethylamide, 3,4-methylenedioxymethamphetamine, ayahuasca, and escitalopram for depressive symptoms: systematic review and Bayesian network meta-analysis. *Brit. Med. J.* **386**, e078607 (2024).
47. Hieronymus, F., López, E., Werin Sjögren, H. & Lundberg, J. Control group outcomes in trials of psilocybin, SSRIs, or esketamine for depression: a meta-analysis. *JAMA Netw. Open* **8**, e2524119 (2025).
48. Goodwin, G. M., Malievskaia, E., Fonzo, G. A. & Nemeroff, C. B. Must psilocybin always 'assist psychotherapy'? *Am. J. Psychiatry* **181**, 20–25 (2024).
49. O'Donnell, K. C. et al. Misinterpretations and omissions: a critical response to goodwin and colleagues' commentary on psilocybin-assisted therapy. *Am. J. Psychiatry* **181**, 74–75 (2024).

50. Bogenschutz, M. P. Pharmacological and nonpharmacological components of psychedelic treatments: the whole is not the sum of the parts. *Am. J. Psychiatry* **181**, 75 (2024).
51. Schenberg, E. E., King, F., da Fonseca, J. E. & Roseman, L. Is poorly assisted psilocybin treatment an increasing risk? *Am. J. Psychiatry* **181**, 75–76 (2024).
52. Alpert, M. D. et al. Psychotherapy in psychedelic treatment: safe, evidence-based, and necessary. *Am. J. Psychiatry* **181**, 76–77 (2024).
53. Deckel, G. M., Lepow, L. A. & Guss, J. 'Psychedelic assisted therapy' must not be retired. *Am. J. Psychiatry* **181**, 77–78 (2024).
54. Earleywine, M., De Leo, J., Bhayana, D., Rajanna, B. & Scott, K. Psilocybin without psychotherapy: a cart without a horse?. *Am. J. Psychiatry* **181**, 78–79 (2024).
55. Goodwin, G. M., Malievskaia, E., Fonzo, G. A. & Nemeroff, C. B. Psychological support for psilocybin treatment: reply to letters on our commentary. *Am. J. Psychiatry* **181**, 79–81 (2024).
56. Higgins, J. P. T. et al. (eds) *Cochrane Handbook for Systematic Reviews of Interventions* Version 6.5 (Cochrane, 2024); https://www.cochrane.org/handbook
57. Viechtbauer, W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J. Educ. Behav. Stat.* **30**, 261–293 (2005).
58. Viechtbauer, W. Confidence intervals for the amount of heterogeneity in meta-analysis. *Stat. Med.* **26**, 37–52 (2007).
59. Knapp, G. & Hartung, J. Improved tests for a random effects meta-regression with a single covariate. *Stat. Med.* **22**, 2693–2710 (2003).
60. Egger, M., Smith, G. D., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *Brit. Med. J.* **315**, 629–634 (1997).
61. Pustejovsky, J. E. & Tipton, E. Meta-analysis with robust variance estimation: expanding the range of working models. *Prev. Sci.* **23**, 425–438 (2022).
62. Paule, R. C. & Mandel, J. Consensus values and weighting factors. *J. Res. Natl Bur. Stand.* **87**, 377 (1982).
63. Williams, D. R., Rast, P. & Bürkner, P. C. Bayesian meta-analysis with weakly informative prior distributions. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/7tbrm (2018).
64. Röver, C. et al. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Res. Synth. Methods* **12**, 448–474 (2021).
65. Harrer, M., Kuper, P., Sprenger, A. & Cuijpers, P. metapsyTools: Several R helper functions for the 'Metapsy' Database. https://tools.metapsy.org (2022).
66. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a practical tutorial. *Evid. Based Ment. Health* **22**, 153–160 (2019).
67. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* https://doi.org/10.18637/jss.v036.i03 (2010).
68. Harrer, M., Cuijpers, P., Furukawa, T. & Ebert, D. D. Dmetar: companion R package for the guide 'Doing Meta-Analysis in R'. http://dmetar.protectlab.org (2019).
69. Röver, C. Bayesian random-effects meta-analysis using the bayesmeta R package. *J. Stat. Softw.* **93**, 1–51 (2020).
70. Singleton, S. P. et al. metapsy-project/data-depression-psiloctr: Version 25.1.2. *Zenodo* https://doi.org/10.5281/zenodo.15714852 (2025).

Acknowledgements

This effort is supported by the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks and Pediatric Anesthesia Safety Initiative (ACTTION/PASI) public–private partnership with the US FDA. The views expressed in this article are those of the authors, and no endorsement by the FDA should be inferred. R.H.D. is director and chair of the ACTTION management committee, and T.D.S. obtained funding for this study via ACTTION.

We thank J. Lege-Matsuura and K. McShea for their assistance with generating database search terms. We thank N. Sepeda, R. von Rotz, J. Rosenblat, N. Rieser, R. Carhart-Harris, B. Szigeti, D. Erritzoe, A. Luquiens, A. Inamdar and S. Kulikova for providing data or other clarifications on studies included in this review.

Author contributions

S.P.S., B.L.S., S.M.N., E.C.S., S.N.V., R.H.D., J.C.S. and T.D.S. conceptualized and pre-registered the study. S.P.S., B.L.S. and A.L. performed literature screening and data extraction. S.P.S. and B.L.S. performed the analysis. S.N.V., M.T.J. and M.H. provided statistical support. M.H. and P.C. provided administrative support for the living database and meta-analytic dashboard via Metapsy. S.P.S., B.L.S., J.C.S. and T.D.S. drafted the original manuscript. T.D.S. obtained funding. T.D.S. and J.C.S. provided supervision. All authors reviewed the manuscript, provided feedback and approved the final version.

Competing interests

Over the past 3 years, E.C.S. has had grant funding to his institution from NIH; editing payments from Wolters–Kluwer; consulting fees from Eli Lilly and Company; medical devices supplied to his institution for his research by Masimo; and he has conducted medical–legal consultations. He also has served on the board of directors (unpaid) for a treatment program, Ashley Addiction Treatment. S.M.N. is a co-investigator on a Usona Institute sponsored trial of psilocybin for major depressive disorder. R.H.D. is the director and chair of the ACTTION management committee that funded this study and, since 1 January 2021, has received research grants and contracts from the US FDA and the US NIH, and compensation for serving on advisory boards or consulting on clinical trial methods from Acadia, Akigai, Allay, AM-Pharma, Analgesic Solutions, Beckley, Biogen, Biosplice, Bsense, Cardialen, Chiesi, Clexio, Collegium, CombiGene, Confo, Contineum, Eccogene, Editas, Eli Lilly, Emmes, Endo, Epizon, Ethismos (equity), Exicure, GlaxoSmithKline, Glenmark, Gloriana, JucaBio, Kriya, Mainstay, Merck, Mind Medicine (also equity), NeuroBo, Noema, OliPass, Orion, Oxford Cannabinoid Technologies, Pfizer, Q-State, Regenacy (also equity), Rho, Salvia, Sangamo, Semnur, SIMR Biotech, Sinfonia, SK Biopharmaceuticals, Sparian, SPM Therapeutics, SPRIM Health, Tiefenbacher, Validae, Vertex, Viscera and WCG. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44220-026-00630-8.

Correspondence and requests for materials should be addressed to S. Parker Singleton or Theodore D. Satterthwaite.

Peer review information *Nature Mental Health* thanks Joshua Black, Yan Liu and Richard J. Zeifman for their contribution to the peer review of this work. Peer reviewer reports are available.

[1]Penn Lifespan Informatics and Neuroimaging Center (PennLINC), Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [2]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [3]Department of Clinical, Neuro and Developmental Psychology, WHO Collaborating Centre for Research and Dissemination of Psychological Interventions, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [4]Babeș-Bolyai University, Ciuj-Napoca, Romania. [5]Amsterdam Public Health Research Institute, Amsterdam University Medical Centre, Amsterdam, the Netherlands. [6]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA. [7]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [8]Departments of Anesthesiology and Perioperative Medicine and Neurology, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA. [9]Department of Psychiatry, Brain Behavior Laboratory, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [10]VISN4 Mental Illness Research, Education and Clinical Center, Crescenz VA Medical Center, Philadelphia, PA, USA. [11]Penn-CHOP Lifespan Brain Institute, Perelman School of Medicine, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA, USA. [12]These authors contributed equally: S. Parker Singleton, Brooke L. Sevchik. [13]These authors jointly supervised this work: J. Cobb Scott, Theodore D. Satterthwaite. ✉e-mail: parker.singleton@pennmedicine.upenn.edu; sattertt@pennmedicine.upenn.edu

nature portfolio

Corresponding author(s): S. Parker Singleton

Last updated by author(s): 03/02/2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Literature screening and data extraction was performed using Covidence.
Data analysis	Meta-analyses were conducted using R (4.4.1) in RStudio (2024.04.2+764) with metapsyTools (1.0.12), a package of helper functions for Metapsy that uses meta (7.0.0) , metafor (4.6.0), and dmetar (0.1.0) functions. The R package bayesmeta (3.5) was used for Bayesian analyses.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The living database used for this analysis can be accessed through Metapsy (docs.metapsy.org/databases/depression-psiloctr/) or downloaded from our website, where code for all analyses is also available (sypres.io).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	While we did not have individual patient data to enable comparison of results by sex or gender, we do provide the % female in each study population, as reported by each study, in Table 1.
Reporting on race, ethnicity, or other socially relevant groupings	We did not have individual patient data to enable these comparisons.
Population characteristics	We report study population age in Table 1. We performed sensitivity analyses on patient diagnoses and co-morbidities with depression.
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	After searching PubMed, Embase, PsycInfo, Web of Science, Scopus, and the reference lists of systematic reviews retrieved from the searches, we identified 8,174 reports and screened them according to pre-specified inclusion criteria. 15 randomized controlled trials comprising 801 participants are included in the database. This comprises all randomized controlled trials fitting our inclusion criteria and is sufficient for a meta-analysis.
Data exclusions	12 of those studies included in our primary model (n = 585). The remaining three studies did not fit the pre-specified inclusion criteria for our primary analysis. The three studies included gray literature (k = 1; Krempien 2023), studies presenting only a mixture of pre- and post-crossover data (k = 1; Grob 2011), and studies comparing psilocybin with comparators with known therapeutic efficacy (k = 1; Carhart-Harris 2021).
Replication	To assess the robustness of our primary findings and explore potential sources of heterogeneity, we conducted a series of five subgroup-specific analyses for continuous outcomes. These analyses allowed us to evaluate the consistency of treatment effects across different study characteristics and methodological approaches: Diagnosis of major depressive disorder (MDD): To assess whether participant diagnosis impacted treatment effects, we conducted a subgroup analysis limited to studies focused on patients with a formal diagnosis of MDD (includes Davis 2021, Goodwin 2022, Raison 2023, von Rotz 2023, and Rosenblat 2024). Exclude open-label: We evaluated the impact of excluding open-label studies (excluding Davis, 2021, Rosenblat, 2024, and Lewis 2025). Exclude high RoB: To assess the impact of study quality on outcomes, we excluded studies rated as having high risk of bias (excluding Griffiths 2016, Rosenblat 2024, and Luquiens 2025). Parallel design: We conducted a subgroup-specific analysis that only included parallel group studies (including Goodwin 2022, Raison 2023, von Rotz 2023, Back 2024, Lewis 2025, Luquiens 2025, and Rieser 2025). Crossover or open-label extension design: We conducted an analysis that only included crossover design studies and studies with open-label extension periods (including Griffiths 2016, Ross 2016, Davis 2021, Rosenblat 2024, and Ross 2025). The following sensitivity analyses were also performed: Expanded inclusion criteria: We conducted an analysis with expanded eligibility criteria that incorporated all studies from the primary model plus three additional studies that were excluded from primary analyses (see Eligibility Criteria). Excluding outliers: We repeated our primary meta-analysis on continuous outcomes after removing statistical outlier studies (e.g., whose effect size confidence intervals do not overlap with the confidence interval of the pooled effect; Davis 2021). Fixed effects model: We ran fixed-effects models as sensitivity analyses to compare to random effects models. Fixed effects models assume that the between-study variance (tau2) is 0, such that all studies share a common true effect size. For our continuous model, we used a standard inverse-variance weighting fixed-effects model on standardized mean differences (Hedges' g). For our dichotomous model, we used a standard inverse-variance weighting fixed-effects model on the log risk ratio. Alternate dosing in Goodwin 2022: Given that Goodwin 2022 employed a three-arm design comparing psilocybin at 25 mg and 10 mg doses against a 1 mg control, we conducted a subgroup-specific analysis substituting the 10 mg intervention arm for the 25 mg arm used in our primary analysis.

Clinician-rated outcomes: To examine whether the method of assessment influenced observed effects, we conducted a sensitivity analysis that only included clinician-administered depression assessments. MADRS was chosen as the preferred clinician-administered instrument for this analysis, followed by the GRID-HAM-D. Studies included were: Goodwin 2022, Raison 2023, von Rotz 2023, Back 2024, and Rosenblat 2024 using the MADRS; and Griffiths 2016 and Davis 2021 using the GRID-HAM-D.
Self-report outcomes: Similarly, we conducted a sensitivity analysis that only included studies reporting self-report depression measures, regardless of whether these were the primary outcome measures. BDI was the preferred self-report instrument which was reported in all self-report studies, except two for which only the QIDS-SR was available. Studies included were: Ross 2016, Griffiths 2016, Davis 2021, Goodwin 2022, von Rotz 2023, Lewis 2025, Luquiens 2025, Rieser 2025, and Ross 2025.
Bayesian meta-analysis: We replicated our primary meta-analysis on continuous outcomes using a Bayesian implementation. We used "weakly informative" prior distributions for both the main effect and the heterogeneity parameter tau that have been recommended by prior work25,26. The main effect prior was a normal distribution centered around 0, with a standard deviation of 1, while the tau prior was a half-normal distribution with a standard deviation of 0.5. This same approach was used in the post-hoc Bayesian analysis of the MDD subgroup.
Self-report outcomes: Similarly, we conducted a sensitivity analysis that only included studies reporting self-report depression measures, regardless of whether these were the primary outcome measures. BDI was the preferred self-report instrument which was reported in all self-report studies. Studies included were: Ross 2016, Griffiths 2016, Davis 2021, von Rotz 2023, and Rieser 2025.
All sub-group and sensitivity analyses produced results in line with our primary analysis with the exception of the MDD subgroup, which was trending towards significance (p = 0.051).

Randomization: The studies we extracted data from were all randomized controlled trials.

Blinding: Blinding was not possible, as we collected data from previously published reports.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks: *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes: *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication: *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*