



# Efficacy of psilocybin for treating symptoms of depression: systematic review and meta-analysis

Athina-Marina Metaxa,<sup>1</sup> Mike Clarke<sup>2</sup>

<sup>1</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford OX2 6GG, UK

<sup>2</sup>Northern Ireland Methodology Hub, Centre for Public Health, ICS-A Royal Hospitals, Belfast, Ireland, UK

Correspondence to: A-M Metaxa [athina.metaxa@hmc.ox.ac.uk](mailto:athina.metaxa@hmc.ox.ac.uk) (or @Athina\_Metaxa12 on X; ORCID 0000-0002-4256-5486) Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2024;385:e078084 <http://dx.doi.org/10.1136/bmj-2023-078084>

Accepted: 06 March 2024

## ABSTRACT

### OBJECTIVE

To determine the efficacy of psilocybin as an antidepressant compared with placebo or non-psychoactive drugs.

### DESIGN

Systematic review and meta-analysis.

### DATA SOURCES

Five electronic databases of published literature (Cochrane Central Register of Controlled Trials, Medline, Embase, Science Citation Index and Conference Proceedings Citation Index, and PsycInfo) and four databases of unpublished and international literature (ClinicalTrials.gov, WHO International Clinical Trials Registry Platform, ProQuest Dissertations and Theses Global, and PsycEXTRA), and handsearching of reference lists, conference proceedings, and abstracts.

### DATA SYNTHESIS AND STUDY QUALITY

Information on potential treatment effect moderators was extracted, including depression type (primary or secondary), previous use of psychedelics, psilocybin dosage, type of outcome measure (clinician rated or self-reported), and personal characteristics (eg, age, sex). Data were synthesised using a random effects meta-analysis model, and observed heterogeneity and the effect of covariates were investigated with subgroup analyses and metaregression. Hedges' *g* was used as a measure of treatment effect size, to account for small sample effects and substantial differences between the included studies' sample sizes. Study quality was appraised using Cochrane's

Risk of Bias 2 tool, and the quality of the aggregated evidence was evaluated using GRADE guidelines.

### ELIGIBILITY CRITERIA

Randomised trials in which psilocybin was administered as a standalone treatment for adults with clinically significant symptoms of depression and change in symptoms was measured using a validated clinician rated or self-report scale. Studies with directive psychotherapy were included if the psychotherapeutic component was present in both experimental and control conditions. Participants with depression regardless of comorbidities (eg, cancer) were eligible.

### RESULTS

Meta-analysis on 436 participants (228 female participants), average age 36-60 years, from seven of the nine included studies showed a significant benefit of psilocybin (Hedges' *g*=0.66, 95% confidence interval (CI) 0.46 to 0.86, *P*<0.001) on change in depression scores compared with comparator treatment. Exploratory subgroup analyses and metaregressions indicated that having secondary depression (Hedges' *g*=0.88, 95% CI 0.42 to 1.33), being assessed with self-report depression scales such as the Beck depression inventory (0.88, 0.42 to 1.33), and older age and previous use of psychedelics (metaregression coefficient 0.13, 95% CI 0.02 to 0.23 and 6.00, 2.48 to 9.53, respectively) were correlated with greater improvements in symptoms. All studies had a moderate risk of bias, but the change from baseline metric was associated with low heterogeneity and a statistically non-significant risk of small study bias, resulting in a moderate certainty of evidence rating.

### CONCLUSION

Treatment effects of psilocybin were significantly larger among patients with secondary depression, when self-report scales were used to measure symptoms of depression, and when participants had previously used psychedelics. Further research is thus required to delineate the influence of expectancy effects, moderating factors, and treatment delivery on the efficacy of psilocybin as an antidepressant.

### SYSTEMATIC REVIEW REGISTRATION

PROSPERO CRD42023388065.

### Introduction

Depression affects an estimated 300 million people around the world, an increase of nearly 20% over the past decade.<sup>1</sup> Worldwide, depression is also the leading cause of disability.<sup>2</sup>

Drugs for depression are widely available but these seem to have limited efficacy, can have serious adverse effects, and are associated with low patient adherence.<sup>3,4</sup> Importantly, the treatment effects of antidepressant drugs do not appear until 4-7 weeks after the start of treatment, and remission of symptoms can take

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Recent research on treatments for depression has focused on psychedelic agents that could have strong antidepressant effects without the drawbacks of classic antidepressants; psilocybin being one such substance

Over the past decade, several clinical trials, meta-analyses, and systematic reviews have investigated the use of psilocybin for symptoms of depression, and most have found that psilocybin can have antidepressant effects

Studies published to date have not investigated factors that may moderate psilocybin's effects, including type of depression, past use of psychedelics, dosage, outcome measures, and publication biases

## WHAT THIS STUDY ADDS

This review showed a significantly greater efficacy of psilocybin among patients with secondary depression, patients with past use of psychedelics, older patients, and studies using self-report measures for symptoms of depression Efficacy did not appear to be homogeneous across patient types—for example, those with depression and a life threatening illness appeared to benefit more from treatment

Further research is needed to clarify the factors that maximise psilocybin's treatment potential for symptoms of depression

months.<sup>4,5</sup> Additionally, the likelihood of relapse is high, with 40-60% of people with depression experiencing a further depressive episode, and the chance of relapse increasing with each subsequent episode.<sup>6,7</sup>

Since the early 2000s, the naturally occurring serotonergic hallucinogen psilocybin, found in several species of mushrooms, has been widely discussed as a potential treatment for depression.<sup>8,9</sup> Psilocybin's mechanism of action differs from that of classic selective serotonin reuptake inhibitors (SSRIs) and might improve the treatment response rate, decrease time to improvement of symptoms, and prevent relapse post-remission. Moreover, more recent assessments of harm have consistently reported that psilocybin generally has low addictive potential and toxicity and that it can be administered safely under clinical supervision.<sup>10</sup>

The renewed interest in psilocybin's antidepressive effects led to several clinical trials on treatment resistant depression,<sup>11,12</sup> major depressive disorder,<sup>13</sup> and depression related to physical illness.<sup>14-17</sup> These trials mostly reported positive efficacy findings, showing reductions in symptoms of depression within a few hours to a few days after one dose or two doses of psilocybin.<sup>11-13,16-18</sup> These studies reported only minimal adverse effects, however, and drug harm assessments in healthy volunteers indicated that psilocybin does not induce physiological toxicity, is not addictive, and does not lead to withdrawal.<sup>19,20</sup> Nevertheless, these findings should be interpreted with caution owing to the small sample sizes and open label design of some of these studies.<sup>11,21</sup>

Several systematic reviews and meta-analyses since the early 2000s have investigated the use of psilocybin to treat symptoms of depression. Most found encouraging results, but as well as people with depression some included healthy volunteers,<sup>22</sup> and most combined data from studies of multiple serotonergic psychedelics,<sup>23-25</sup> even though each compound has unique neurobiological effects and mechanisms of action.<sup>26-28</sup> Furthermore, many systematic reviews included non-randomised studies and studies in which psilocybin was tested in conjunction with psychotherapeutic interventions,<sup>25,29-32</sup> which made it difficult to distinguish psilocybin's treatment effects. Most systematic reviews and meta-analyses did not consider the impact of factors that could act as moderators to psilocybin's effects, such as type of depression (primary or secondary), previous use of psychedelics, psilocybin dosage, type of outcome measure (clinician rated or self-reported), and personal characteristics (eg, age, sex).<sup>25,26,29-32</sup> Lastly, systematic reviews did not consider grey literature,<sup>33,34</sup> which might have led to a substantial overestimation of psilocybin's efficacy as a treatment for depression. In this review we focused on randomised trials that contained an unconfounded evaluation of psilocybin in adults with symptoms of depression, regardless of country and language of publication.

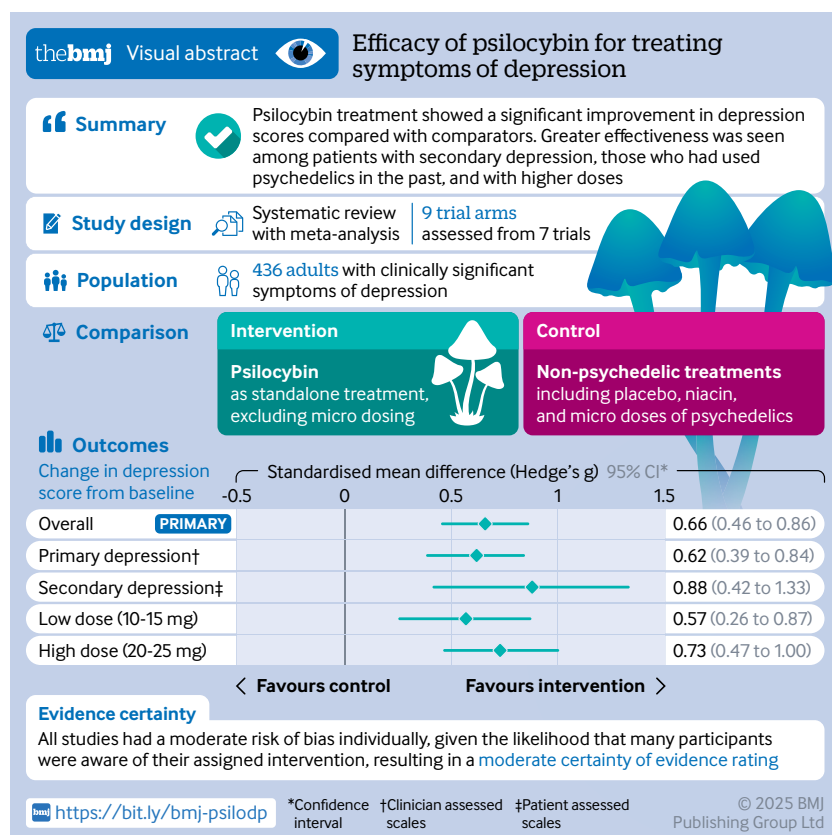
## Methods

In this systematic review and meta-analysis of indexed and non-indexed randomised trials we investigated the efficacy of psilocybin to treat symptoms of depression compared with placebo or non-psychoactive drugs. The protocol was registered in the International Prospective Register of Systematic Reviews (see supplementary Appendix A). The study overall did not deviate from the pre-registered protocol; one clarification was made to highlight that any non-psychedelic comparator was eligible for inclusion, including placebo, niacin, micro doses of psychedelics, and drugs that are considered the standard of care in depression (eg, SSRIs).

## Inclusion and exclusion criteria

Double blind and open label randomised trials with a crossover or parallel design were eligible for inclusion. We considered only studies in humans and with a control condition, which could include any type of non-active comparator, such as placebo, niacin, or micro doses of psychedelics.

Eligible studies were those that included adults ( $\geq 18$  years) with clinically significant symptoms of depression, evaluated using a clinically validated tool for depression and mood disorder outcomes. Such tools included the Beck depression inventory, Hamilton depression rating scale, Montgomery-Åsberg depression rating scale, profile of mood states, and quick inventory of depressive symptomatology. Studies of participants with symptoms of depression and comorbidities (eg, cancer) were also eligible. We



excluded studies of healthy participants (without depressive symptomatology).

Eligible studies investigated the effect of psilocybin as a standalone treatment on symptoms of depression. Studies with an active psilocybin condition that involved micro dosing (ie, psilocybin <100 µg/kg, according to the commonly accepted convention<sup>22,35</sup>) were excluded. We included studies with directive psychotherapy if the psychotherapeutic component was present in both the experimental and the control conditions, so that the effects of psilocybin could be distinguished from those of psychotherapy. Studies involving group therapy were also excluded. Any non-psychedelic comparator was eligible for inclusion, including placebo, niacin, and micro doses of psychedelics.

Changes in symptoms, measured by validated clinician rated or self-report scales, such as the Beck depression inventory, Hamilton depression rating scale, Montgomery-Åsberg depression rating scale, profile of mood states, and quick inventory of depressive symptomatology were considered. We excluded outcomes that were measured less than three hours after psilocybin had been administered because any reported changes could be attributed to the transient cognitive and affective effects of the substance being administered. Aside from this, outcomes were included irrespective of the time point at which measurements were taken.

### Search strategy

We searched major electronic databases and trial registries of psychological and medical research, with no limits on the publication date. Databases were the Cochrane Central Register of Controlled Trials via the Cochrane Library, Embase via Ovid, Medline via Ovid, Science Citation Index and Conference Proceedings Citation Index-Science via Web of Science, and PsycInfo via Ovid. A search through multiple databases was necessary because each database includes unique journals. Supplementary Appendix B shows the search syntax used for the Cochrane Central Register of Controlled Trials, which was slightly modified to comply with the syntactic rules of the other databases.

Unpublished and grey literature were sought through registries of past and ongoing trials, databases of conference proceedings, government reports, theses, dissertations, and grant registries (eg, ClinicalTrials.gov, WHO International Clinical Trials Registry Platform, ProQuest Dissertations and Theses Global, and PsycEXTRA). The references and bibliographies of eligible studies were checked for relevant publications. The original search was done in January 2023 and updated search was performed on 10 August 2023.

### Data collection, extraction, and management

The results of the literature search were imported to the Endnote X9 reference management software, and the references were imported to the Covidence platform after removal of duplicates. Two reviewers (AM and DT) independently screened the title and abstract of each reference and then screened the full text of potentially eligible references. Any disagreements about eligibility

were resolved through discussion. If information was insufficient to determine eligibility, the study's authors were contacted. The reviewers were not blinded to the studies' authors, institutions, or journal of publication.

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram shows the study selection process and reasons for excluding studies that were considered eligible for full text screening.<sup>36</sup>

### Critical appraisal of individual studies and of aggregated evidence

The methodological quality of eligible studies was assessed using the Cochrane Risk of Bias 2 tool (RoB 2) for assessing risk of bias in randomised trials.<sup>37</sup> In addition to the criteria specified by RoB 2, we considered the potential impact of industry funding and conflicts of interest. The overall methodological quality of the aggregated evidence was evaluated using GRADE (Grading of Recommendations, Assessment, Development and Evaluation).<sup>38</sup>

If we found evidence of heterogeneity among the trials, then small study biases, such as publication bias, were assessed using a funnel plot and asymmetry tests (eg, Egger's test).<sup>39</sup>

### Data items

We used a template for data extraction (see supplementary Appendix C) and summarised the extracted data in tabular form, outlining personal characteristics (age, sex, previous use of psychedelics), methodology (study design, dosage), and outcome related characteristics (mean change from baseline score on a depression questionnaire, response rates, and remission rates) of the included studies. Response conventionally refers to a 50% decrease in symptom severity based on scores on a depression rating scale, whereas remission scores are specific to a questionnaire (eg, score of ≤5 on the quick inventory of depressive symptomatology, score of ≤10 on the Montgomery-Åsberg depression rating scale, 50% or greater reduction in symptoms, score of ≤7 on the Hamilton depression rating scale, or score of ≤12 on the Beck depression inventory). Across depression scales, higher scores signify more severe symptoms of depression.

### Continuous data synthesis

From each study we extracted the baseline and post-intervention means and standard deviations (SDs) of the scores between comparison groups for the depression questionnaires and calculated the mean differences and SDs of change. If means and SDs were not available for the included studies, we extracted the values from available graphs and charts using the Web Plot Digitizer application (<https://automeris.io/WebPlotDigitizer/>). If it was not possible to calculate SDs from the graphs or charts, we generated values by converting standard errors (SEs) or confidence intervals (CIs), depending on availability, using formulas in the Cochrane Handbook (section 7.7.3.2).<sup>40</sup>



Standardised mean differences were calculated for each study. We chose these rather than weighted mean differences because, although all the studies measured depression as the primary outcome, they did so with different questionnaires that score depression based on slightly different items.<sup>41</sup> If we had used weighted mean differences, any variability among studies would be assumed to reflect actual methodological or population differences and not differences in how the outcome was measured, which could be misleading.<sup>40</sup>

The Hedges' *g* effect size estimate was used because it tends to produce less biased results for studies with smaller samples (<20 participants) and when sample sizes differ substantially between studies, in contrast with Cohen's *d*.<sup>42</sup> According to the Cochrane Handbook, the Hedges' *g* effect size measure is synonymous with the standardised mean difference,<sup>40</sup> and the terms may be used interchangeably. Thus, a Hedges' *g* of 0.2, 0.5, 0.8, or 1.2 corresponds to a small, medium, large, or very large effect, respectively.<sup>40</sup>

Owing to variation in the participants' personal characteristics, psilocybin dosage, type of depression investigated (primary or secondary), and type of comparators, we used a random effects model with a Hartung-Knapp-Sidik-Jonkman modification.<sup>43</sup> This model also allowed for heterogeneity and within study variability to be incorporated into the weighting of the results of the included studies.<sup>44</sup> Lastly, this model could help to generalise the findings beyond the studies and patient populations included, making the meta-analysis more clinically useful.<sup>45</sup> We chose the Hartung-Knapp-Sidik-Jonkman adjustment in favour of more widely used random effects models (eg, DerSimonian and Laird) because it allows for better control of type 1 errors, especially for studies with smaller samples, and provides a better estimation of between study variance by accounting for small sample sizes.<sup>46 47</sup>

For studies in which multiple treatment groups were compared with a single placebo group, we split the placebo group to avoid multiplicity.<sup>48</sup> Similarly, if studies included multiple primary outcomes (eg, HADS-D and BDI), we split the treatment groups to account for overlapping participants.<sup>40</sup>

Prediction intervals (PIs) were calculated and reported to show the expected effect range of a similar future study, in a different setting. In a random effects model, within study measures of variability, such as CIs, can only show the range in which the average effect size could lie, but they are not informative about the range of potential treatment effects given the heterogeneity between studies.<sup>49</sup> Thus, we used PIs as an indication of variation between studies.

### Heterogeneity and sensitivity analysis

Statistical heterogeneity was tested using the  $\chi^2$  test (significance level  $P < 0.1$ ) and  $I^2$  statistic, and heterogeneity among included studies was evaluated visually and displayed graphically using a forest plot. If substantial or considerable heterogeneity was found ( $I^2 \geq 50\%$  or  $P < 0.1$ ),<sup>50</sup> we considered the study design and characteristics of the included studies. Sources

of heterogeneity were explored by subgroup analysis, and the potential effects on the results are discussed.

Planned sensitivity analyses to assess the effect of unpublished studies and studies at high risk of bias were not done because all included studies had been published and none were assessed as high risk of bias. Exclusion sensitivity plots were used to display graphically the impact of individual studies and to determine which studies had a particularly large influence on the results of the meta-analysis. All sensitivity analyses were carried out with Stata 16 software.

### Subgroup analysis

To reduce the risk of errors caused by multiplicity and to avoid data fishing, we planned subgroup analyses a priori and limited to: (1) patient characteristics, including age and sex; (2) comorbidities, such as a serious physical condition (previous research indicates that the effects of psilocybin may be less strong for such participants, compared with participants with no comorbidities)<sup>33</sup>; (3) number of doses and amount of psilocybin administered, because some previous meta-analyses found that a higher number of doses and a higher dose of psilocybin both predicted a greater reduction in symptoms of depression,<sup>34</sup> whereas others reported the opposite<sup>33</sup>; (4) psilocybin administered alongside psychotherapeutic guidance or as a standalone treatment; (5) severity of depressive symptoms (clinical v subclinical symptomatology); (6) clinician versus patient rated scales; and (7) high versus low quality studies, as determined by RoB 2 assessment scores.

### Metaregression

Given that enough studies were identified ( $\geq 10$  distinct observations according to the Cochrane Handbook's suggestion<sup>40</sup>), we performed exploratory metaregressions to investigate whether covariates, or potential effect modifiers, explained any of the statistical heterogeneity. The metaregression analysis was carried out using Stata 16 software.

Random effects metaregression analyses were used to determine whether continuous variables such as participants' age, percentage of female participants, and percentage of participants who had previously used psychedelics modified the effect estimate, all of which have been implicated in differentially affecting the efficacy of psychedelics in modifying mood.<sup>51</sup> We chose this approach in favour of converting these continuous variables into categorical variables and conducting subgroup analyses for two primary reasons; firstly, the loss of any data and subsequent loss of statistical power would increase the risk of spurious significant associations,<sup>51</sup> and, secondly, no cut-offs have been agreed for these factors in literature on psychedelic interventions for mood disorders,<sup>52</sup> making any such divisions arbitrary and difficult to reconcile with the findings of other studies. The analyses were based on within study averages, in the absence of individual data points for each participant, with the potential for the results to be affected by aggregate bias, compromising their validity and

generalisability.<sup>53</sup> Furthermore, a group level analysis may not be able to detect distinct interactions between the effect modifiers and participant subgroups, resulting in ecological bias.<sup>54</sup> As a result, this analysis should be considered exploratory.

### Sensitivity analysis

A sensitivity analysis was performed to determine if choice of analysis method affected the primary findings of meta-analysis. Specifically, we reanalysed the data on change in depression score using a random effects Dersimonian and Laird model without the Hartung-Knapp-Sidik-Jonkman modification and compared the results with those of the originally used model. This comparison is particularly important in the presence of substantial heterogeneity and the potential of small study effects to influence the intervention effect estimate.<sup>55</sup>

### Patient and public involvement

Research on novel depression treatments is of great interest to both patients and the public. Although patients and members of the public were not directly involved in the planning or writing of this manuscript owing to a lack of available funding for recruitment and researcher training, patients and members of the public read the manuscript after submission.

### Results

Figure 1 presents the flow of studies through the systematic review and meta-analysis.<sup>56</sup> A total of 4884 titles were retrieved from the five databases of published literature, and a further 368 titles were identified from the databases of unpublished and international literature in February 2023. After the removal of duplicate records, we screened the abstracts and titles of 875 reports. A further 12 studies were added after handsearching of reference lists and conference proceedings and abstracts. Overall, nine studies totalling 436 participants were eligible. The average age of the participants ranged from 36–60 years. During an updated search on 10 August 2023, no further studies were identified.

After screening of the title and abstract, 61 titles remained for full text review. Native speakers helped to translate papers in languages other than English. The most common reasons for exclusion were the inclusion of healthy volunteers, absence of control groups, and use of a survey based design rather than an experimental design. After full text screening, nine studies were eligible for inclusion, and 15 clinical trials prospectively registered or underway as of August 2023 were noted for potential future inclusion in an update of this review (see supplementary Appendix D).

We sent requests for further information to the authors of studies by Griffiths et al,<sup>57</sup> Barrett,<sup>58</sup> and Benville et al,<sup>59</sup> because these studies appeared to meet the inclusion criteria but were only provided as summary abstracts online. A potentially eligible poster presentation from the 58th annual meeting of the American College of Neuropsychopharmacology was identified but the lead author (Griffiths) clarified that all information from the presentation was included in the studies by Davis et al<sup>13</sup> and Gukasyan et al<sup>60</sup>; both of which we had already deemed ineligible.

Barrett<sup>58</sup> reported the effects of psilocybin on the cognitive flexibility and verbal reasoning of a subset of patients with major depressive disorder from Griffith et al's trial,<sup>61</sup> compared with a waitlist group, but when contacted, Barrett explained that the results were published in the study by Doss et al,<sup>62</sup> which we had already screened and judged ineligible (see supplementary Appendix E). Benville et al's study<sup>59</sup> presented a follow-up of Ross et al's study<sup>17</sup> on a subset of patients with cancer and high suicidal ideation and desire for hastened death at baseline. Measures of antidepressant effects of psilocybin treatment compared with niacin were taken before and after treatment crossover, but detailed results are not reported. Table 1 describes the characteristics of the included studies and table 2 lists the main findings of the studies.

### Side effects and adverse events

Side effects reported in the included studies were minor and transient (eg, short term increases in blood pressure, headache, and anxiety), and none were coded as serious. Cahart-Harris et al noted one instance of abnormal dreams and insomnia.<sup>63</sup> This

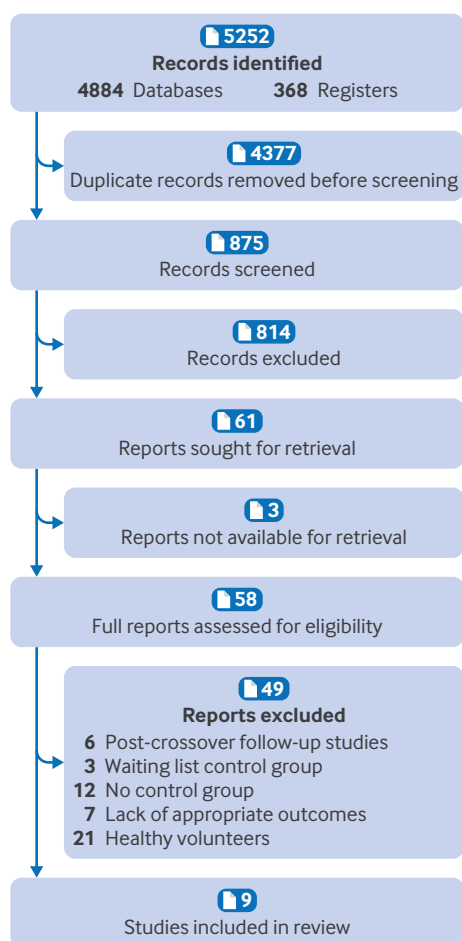


Fig 1 | Flow of studies in systematic review and meta-analysis

Table 1 | Characteristics of included studies

Study, year, reference	Design	Depression type	No of participants	Psilocybin intervention	Control condition	Mean (SD) age (years)	% (No/total No)		Primary outcome or measure of interest
							Female participants	Past use of psychedelics	
Carhart-Harris et al 2021 <sup>63</sup>	Double blind RCT	Primary	59	2 high doses (25 mg) of psilocybin and daily placebo, along with psychological support	2 placebo-like doses of psilocybin (1 mg) and escitalopram, along with psychological support	41.2 (10.7)	34 (20/59)	27.1 (16/59)	QIDS (change from baseline at 6 weeks), patient reported
Barba et al 2022 <sup>64</sup>	Further analysis of Carhart-Harris et al, 2021	Primary	59	2 high doses (25 mg) of psilocybin and daily placebo, along with psychological support	2 placebo-like doses of psilocybin (1 mg) and escitalopram, along with psychological support	41.2 (10.7)	34 (20/59)	27.1 (16/59)	RRS for rumination
Goodwin et al 2022 <sup>18</sup>	Double blind RCT	Primary	233	Moderate dose (10 mg) or high dose (25 mg) of psilocybin and psychological support	Placebo-like dose of psilocybin (1 mg) and psychological support	39.8 (12.2)	52 (121/233)	6.0 (14/233)	MADRS (change from baseline at 3 weeks), clinician assessed
Goodwin et al 2023 <sup>65</sup>	Double blind RCT (further analysis of Goodwin et al, 2023)	Primary	233	Moderate dose (10 mg) or high dose (25 mg) of psilocybin and psychological support	Placebo-like dose of psilocybin (1 mg) and psychological support	39.8 (12.2)	52 (121/233)	6.0 (14/233)	QIDS (change from baseline at 3 weeks), patient reported
von Rotz et al 2023 <sup>66</sup>	Double blind RCT	Primary	52	0.215 mg/kg psilocybin and psychological support	Placebo and psychological support	36.8 (10.4)	64 (33/52)	30.8 (16/52)	MADRS (change from baseline at 2 weeks), clinician assessed
Grob et al 2011 <sup>15</sup>	Randomised, double blind, crossover trial	Secondary (patients with life threatening illness)	12	Moderate dose (0.2 mg/kg) of psilocybin	Placebo, niacin (250 mg)	36-58 (mean not given)	92 (11/12)	66.7 (8/12)	BDI (change from baseline at 2 weeks), patient reported
Griffiths et al 2016 <sup>14</sup>	Randomised, double blind, crossover trial	Secondary (patients with life threatening illness)	51	High dose (22 mg/70 kg or 30 mg/70 kg) of psilocybin	Very low (placebo-like) psilocybin dose (1 mg or 3 mg/70 kg)	56.3 (1.4)	49 (25/51)	45.0 (23/51)	BDI (change from baseline at 5 weeks), patient reported
Ross et al 2016 <sup>17</sup>	Double blind, placebo controlled, crossover trial	Secondary (patients with life threatening illness)	29	Single dose psilocybin (0.3 mg/kg) and psychotherapy	Placebo, niacin (250 mg), and psychotherapy	56.28 (12.9)	62 (18/29)	55.2 (16/29)	BDI (change from baseline at 2 and 6 weeks), patient reported
Ross et al 2021 <sup>67</sup>	Secondary analysis of Ross et al 2016	Secondary (patients with life threatening illness)	11	Single dose psilocybin (0.3 mg/kg) and psychotherapy	Placebo, niacin (250 mg), and psychotherapy	60.3 (7.1)	7/11	4/11	Demoralisation scale

BDI=Beck depression inventory; MADRS=Montgomery-Åsberg depression rating scale; QIDS=quick inventory of depressive symptomatology; RRS=ruminative response scale; RCT=randomised controlled trial; SD=standard deviation.

side effect profile is consistent with findings from other meta-analyses.<sup>30 68</sup> Owing to the different scales and methods used to catalogue side effects and adverse events across trials, it was not possible to combine these data quantitatively (see supplementary Appendix F).

### Risk of bias

The Cochrane RoB 2 tools were used to evaluate the included studies (table 3). RoB 2 for randomised trials was used for the five reports of parallel randomised trials (Carhart-Harris et al<sup>63</sup> and its secondary analysis Barba et al,<sup>64</sup> Goodwin et al<sup>18</sup> and its secondary analysis Goodwin et al,<sup>65</sup> and von Rotz et al<sup>66</sup>) and RoB 2 for crossover trials was used for the four reports of crossover randomised trials (Griffiths et al,<sup>14</sup> Grob et al,<sup>15</sup> and Ross et al<sup>17</sup> and its follow-up Ross et al<sup>67</sup>). Supplementary Appendix G provides a detailed explanation of the assessment of the included studies.

### Quality of included studies

Confidence in the quality of the evidence for the meta-analysis was assessed using GRADE,<sup>38</sup> through the GRADEpro GDT software program. Figure 2 shows the

results of this assessment, along with our summary of findings.

### Meta-analyses

**Continuous data, change in depression scores**—Using a Hartung-Knapp-Sidik-Jonkman modified random effects meta-analysis, change in depression scores was significantly greater after treatment with psilocybin compared with active placebo. The overall Hedges' *g* (0.66, 95% CI 0.46 to 0.86) indicated a moderate-to-large effect size favouring psilocybin (fig 3). PIs were relatively narrow and did not cross the line of no difference (95% CI 0.46 to 0.86), indicating that psilocybin interventions' efficacy could be robust and generalisable across different settings and populations.

**Exploring publication bias in continuous data**—We used Egger's test and a funnel plot to examine the possibility of small study biases, such as publication bias. Statistical non-significance of Egger's test for small study effects ( $P>0.05$ ) and minor asymmetries in the corresponding funnel plot (fig 4), indicates no strong evidence for the presence of bias against smaller studies with non-significant results.<sup>69</sup>

Table 2 | Main findings of included studies

Study	Primary outcome or measure of interest	Change in depression scores	Response rate	Remission rate
Carhart-Harris et al 2021 <sup>63</sup>	QIDS (change from baseline at 6 weeks), patient reported	–8 points for psilocybin v –6 points for escitalopram (P>0.05)	21 (70%) for psilocybin v 14 (48%) for escitalopram (non-significant difference: 95% CI –3% to 48%)	17 (57%) for psilocybin v 8 (29%) for escitalopram (95% CI 2.3% to 53.8%)
Barba et al 2022 <sup>64</sup>	RRS for rumination (change from baseline at 6 weeks)	Significant change from baseline for psilocybin (mean difference –7.76, P<0.001) Non-significant change from baseline for escitalopram (P>0.05)	Not reported	Not reported
Goodwin et al 2022 <sup>18</sup>	MADRS (change from baseline at 3 weeks), clinician assessed	–12.0 points for 25 mg group v –5.4 points for 1 mg group (95% CI –10.2 to –2.9, P<0.001) –7.9 points for 10 mg group v –5.4 points for 1 mg group (P>0.05)	29 (37%) in 25 mg group v 14 (19%) in 10 mg group (odds ratio 2.9, 95% CI 1.2 to 6.6) 14 (19%) in 10 mg group v 14 (18%) in 1 mg group (P>0.05)	23 (29%) in 25 mg group v 7 (9%) in 10 mg group (odds ratio 4.8, 95% CI 1.8 to 12.8) 7 (9%) in 10 mg group v 6 (8%) in 1 mg group (P>0.05)
Goodwin et al 2023 <sup>65</sup>	QIDS (change from baseline at 3 weeks), patient reported	–6.3 points for 25 mg group v –3.6 points for 1 mg group (95% CI –4.6 to –0.9) –5.2 points for 10 mg group v –3.6 points for 1 mg group (P>0.05)	Not reported	Not reported
von Rotz et al 2023 <sup>66</sup>	MADRS (change from baseline at 2 weeks), clinician assessed	–13.0 points in MADRS (95% CI –15.0 to –1.3, P=0.001) –13.2 points in BDI (95% CI –13.4 to –1.3, P=0.02)	MADRS: 15 (58%) for psilocybin v 4 (15%) for placebo (P=0.003) BDI: 14 (54%) for psilocybin v 3 (12%) for placebo (P=0.003)	MADRS: 14 (54%) for psilocybin v 3 (12%) for placebo (P=0.002) BDI: 12 (46%) for psilocybin v 3 (12%) for placebo (P=0.01)
Grob et al 2011 <sup>15</sup>	BDI (change from baseline at 2 weeks), patient reported	No significant changes from baseline in either group (P>0.05)	Not reported	Not reported
Griffiths et al 2016 <sup>14</sup>	BDI (change from baseline at 5 weeks), patient reported	–10.8 points for 22 mg group v –5.5 points for 1 mg group (P<0.01)	92% in 22 mg group v 32% in 1 mg group (P<0.001)	60% in 22 mg group v 16% in 1 mg group (P<0.01)
Ross et al 2016 <sup>17</sup>	BDI (change from baseline at 2 and 6 weeks), patient reported	–8.5 for psilocybin v –3.3 for placebo at 2 weeks (P<0.05) and –8.6 for psilocybin v –2.8 for placebo at 6 weeks (P<0.05)	86% in psilocybin group v 16% for placebo (P<0.01)	86% in psilocybin group v 16% for placebo (P<0.01)
Ross et al 2021 <sup>67</sup>	Demoralisation scale	Significantly lower post-treatment scores for psilocybin (P<0.001), but not placebo (P>0.05)	Not reported	Not reported

BDI=Beck depression inventory; CI=confidence interval; MADRS=Montgomery-Åsberg depression rating scale; QIDS=quick inventory of depressive symptomatology, RRS=ruminative response scale; SD=standard deviation.

### Dichotomous data

We extracted response and remission rates for each group when reported directly, or imputed information when presented graphically. Two studies did not measure response or remission and thus did not contribute data for this part of the analysis.<sup>15 18</sup> The random effects model with a Hartung-Knapp-Sidik-Jonkman modification was used to allow for

heterogeneity to be incorporated into the weighting of the included studies' results, and to provide a better estimation of between study variance accounting for small sample sizes.

**Response rate**—Overall, the likelihood of psilocybin intervention leading to treatment response was about two times greater (risk ratio 2.17, 95% CI 1.18 to 3.99) than with placebo. There was moderate statistical

Table 3 | Summary risk of bias assessment of included studies, based on domains in Cochrane Risk of Bias 2 tool

Study reference	Design	Experimental condition	Comparator	Domain							
				D1	DS	D2	D3	D4	D5	Other	
Carhart-Harris et al 2021, <sup>63</sup> and Barba et al 2022 <sup>64</sup> (secondary analysis)	Parallel	Two doses of 25 mg psilocybin, daily placebo, and psychological support	Two doses of 1 mg psilocybin, daily escitalopram, and psychological support	Low	—	Some concerns	Low	Low	Low	Some concern	
Goodwin et al 2022, <sup>18</sup> and Goodwin et al 2023 <sup>65</sup> (secondary analysis)	Parallel	Single dose (25 mg or 10 mg) psilocybin with psychological support	Placebo-like dose (25 mg or 10 mg) psilocybin with psychological support	Low	—	Some concerns	Low	Low	Low	Some concern	
von Rotz et al 2023 <sup>66</sup>	Parallel	Single dose (0.215 mg/kg) psilocybin with psychological support	Placebo and psychological support	Low	—	Some concerns	Low	Low	Low	Some concern	
Grob et al 2011 <sup>15</sup>	Crossover	Single dose (0.2 mg/kg) psilocybin	Niacin (250 mg)	Low	Low	Some concerns	Low	Low	Low	Some concern	
Griffiths et al 2016 <sup>14</sup>	Crossover	Single dose (22 mg/70 kg or 30 mg/70 kg) psilocybin	Placebo-like (1 mg/70 kg or 3 mg/70 kg) psilocybin	Low	Low	Some concerns	Low	Low	Low	Some concern	
Ross et al 2016, <sup>17</sup> and Ross et al 2021 <sup>67</sup> (secondary analysis)	Crossover	Single dose psilocybin (0.3 mg/kg) and psychotherapy	Niacin (250 mg) and psychotherapy	Low	Low	Some concerns	Low	Low	Low	Some concern	

Domain 1 assessed bias arising from the randomisation process, including blinding and randomisation of the allocation sequence, and baseline differences between groups. Domain 2 assessed bias due to deviations from the intended intervention, including participant and researcher blinding, and the effects of differential adherence between groups. Domain 3 assessed bias due to missing outcome data. Domain 4 assessed bias in measurement of the outcome, including the appropriateness of metrics and questionnaires used and between group differences in outcome assessment. Domain 5 assessed reporting bias arising from the selective reporting of results or data analyses, or both. The "other" criterion assessed bias due to potential conflicts of interest, such as industry funding and associations with pharmaceutical companies.



# Psilocybin compared with placebo for patients with depression

**Patient or population:** Patients with symptoms of depression

**Setting:** Outpatient clinics, practices, or hospitals

**Intervention:** Psilocybin

**Comparison:** Placebo

Outcomes	No of participants (studies)	Certainty of evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with psilocybin
Change in depression scores from baseline (standardised mean difference) assessed with: MADRS, BDI, QIDS	436 (7 RCTs)	⊕⊕○○ Moderate	—	—	Standardised mean difference 0.66 SD higher (0.46 higher to 0.86 higher)
Response to treatment (odds ratio) assessed with: MADRS, QIDS, BDI, HAM-D	424 (6 RCTs)	⊕⊕⊕⊕ High	Risk ratio 2.17 (1.18 to 3.99)	254 per 1000	297 more per 1000 (46 more to 759 more)
Remission (odds ratio) assessed with: MADRS, QIDS, BDI, HAM-D	424 (6 RCTs)	⊕⊕⊕⊕ High	Risk ratio 2.96 (1.69 to 5.17)	145 per 1000	284 more per 1000 (100 more to 605 more)

GRADE Working Group grades of evidence

High certainty: very confident that the true effect lies close to the estimate of the effect

Moderate certainty: moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility it is substantially different

Low certainty: limited confidence in the effect estimate: the true effect may be substantially different from the estimate of the effect

Very low certainty: very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

**Fig 2 | GRADE assessment outputs for outcomes investigated in meta-analysis (change in depression scores and response and remission rates).** The risk in the intervention group (and its 95% CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI). BDI=Beck depression inventory; CI=confidence interval; GRADE=Grading of Recommendations, Assessment, Development and Evaluation; HADS-D=hospital anxiety and depression scale; HAM-D=Hamilton depression rating scale; MADRS=Montgomery-Åsberg depression rating scale; QIDS=quick inventory of depressive symptomatology; RCT=randomised controlled trial; SD=standard deviation

heterogeneity between studies ( $I^2=53.1\%$ ,  $P=0.06$ ). PIs were, however, wide and crossed the line of no difference (0.57 to 2.65), indicating that there could be settings or populations in which psilocybin intervention would be less efficacious.

**Remission rate**—Overall, the likelihood of psilocybin intervention leading to remission of depression was nearly three times greater than with placebo (risk ratio 2.96, 95% CI 1.69 to 5.17). Despite the use of different scales to measure response, there was little statistical heterogeneity between studies ( $I^2=7.1\%$ ,  $P=0.37$ ). PIs marginally crossed the line of no difference (0.93 to 2.38), indicating that there could be settings or populations in which psilocybin intervention would be less efficacious.

**Exploring publication bias in response and remission rates data**—We used Egger's test and a funnel plot to examine whether response and remission estimates were affected by small study biases. The result for Egger's test was non-significant ( $P>0.05$ ) for remission estimates, and no substantial asymmetry was observed in the funnel plots, providing no indication for the presence of bias against smaller studies with non-significant results. Egger's test was significant for response estimates ( $P=0.002$ ), which could indicate

that there might be settings or populations where psilocybin's effects on treatment response are less robust. However, alternative explanations for funnel plot asymmetry are worth considering. For example, it is possible that smaller studies conducted at the early stages of a new psychotherapeutic intervention tend to include more high risk or responsive participants, and psychotherapeutic interventions tend to be delivered more effectively in smaller trials; both of these factors can exaggerate treatment response rates, resulting in some degree of funnel plot asymmetry.<sup>70</sup> However, the relatively small number of included studies means that test power might be insufficient to distinguish real asymmetry from chance.<sup>71</sup> Thus, this analysis should be considered exploratory.

## Heterogeneity: subgroup analyses and metaregression

While heterogeneity was minor across studies for changes in depression scores ( $I^2=0.0\%$ ,  $P>0.05$ ), due to the relatively small number of studies and the fact that multiple arms of single studies contributed to the meta-analyses, subgroup analyses and meta-regression analyses were used to explore possible factors moderating psilocybin's treatment



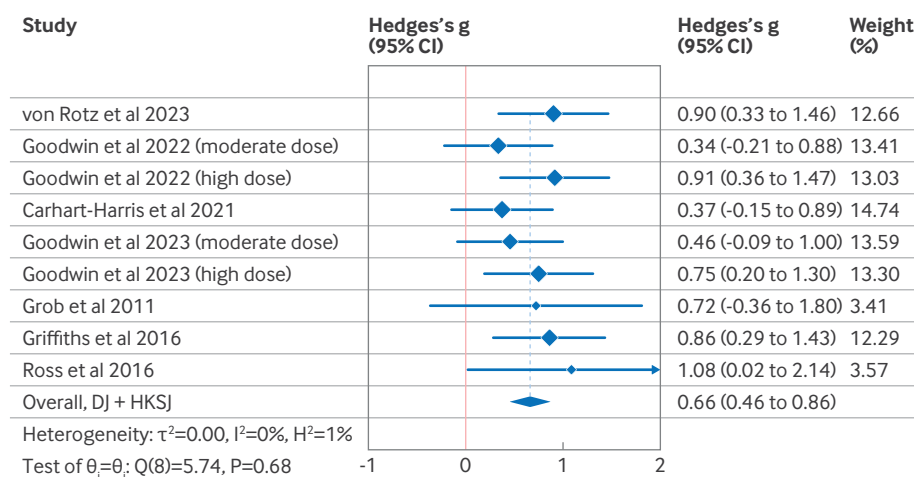


Fig 3 | Forest plot for overall change in depression scores from before to after treatment. CI=confidence interval; DL=DerSimonian and Laird; HKSJ=Hartung-Knapp-Sidik-Jonkman

effectiveness, which have been highlighted across the relevant literature. Given the small number of included studies and the use of graphical extraction software for the extraction of some data values, these analyses should be considered exploratory. While the included studies' authors corroborated the values derived in most cases, this was not possible in the study of Ross et al.<sup>17</sup> Table 4 and table 5 present the results of the heterogeneity analyses (subgroup analyses and meta-regression, respectively).

#### Cumulative meta-analyses

We used cumulative meta-analyses to investigate how the overall estimates of the outcomes of interest changed as each study was added in chronological order<sup>72</sup>; change in depression scores and likelihood of treatment response both increased as the percentage of participants with past use of psychedelics increased across studies, as expected based on the metaregression analysis (see supplementary Appendix I). No other significant time related patterns were found.

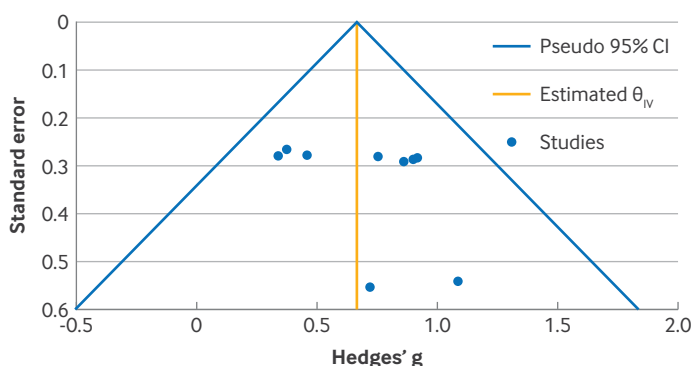


Fig 4 | Funnel plot assessing publication bias among studies measuring change in depression scores from before to after treatment. CI=confidence interval;  $\theta_{IV}$ =estimated effect size under inverse variance random effects model

#### Sensitivity analysis

We reanalysed the data for change in depression scores using a random effects DerSimonian and Laird model without the Hartung-Knapp-Sidik-Jonkman modification and compared the results with those of the original model. Psilocybin's overall treatment benefit was found to be significant when the DerSimonian and Laird model both with and without the Hartung-Knapp-Sidik-Jonkman adjustment was applied, and confidence intervals were only slightly narrower. No significant subgroup effects were found when the DerSimonian and Laird model without the Hartung-Knapp-Sidik-Jonkman was used, but the association between prior use of psychedelics and greater treatment response and remission rates and the association between older age and greater remission rates remained significant. Thus, small study effects do not appear to have played a major role in the treatment effect estimate. Additionally, to estimate the accuracy and robustness of the estimated treatment effect, we excluded studies from the meta-analysis one by one; no important differences in the treatment effect, significance, and heterogeneity levels were observed after the exclusion of any study. To test the effects of our decision to split the sample in Goodwin et al<sup>18</sup> and its follow-up<sup>65</sup> to account for the inclusion of both the MADRS and QIDS outcomes, an analysis was conducted in which only one of these outcomes was used, and thus the sample did not need to be split. In the study of Ross et al,<sup>17</sup> both a two-week and six-week outcome was reported; while the outcome at six weeks was included in the main meta-analysis, because it was deemed more clinically relevant, a sensitivity analysis where the two-week outcome was used instead was also performed.

Across all sensitivity analyses, the main treatment effect of psilocybin and the pattern of subgroup effects seen for depression type and assessment measure remain consistent with those seen in the main analysis. This indicated that the findings are

**Table 4 | Subgroup analyses to explore potential causes of heterogeneity among included studies**

Outcome	Subgroups	Between group heterogeneity	Effect size with confidence intervals and heterogeneity ( $I^2$ ) per subgroup	Conclusion
Change in depression scores	Type of depression: primary or secondary	Significant ( $P=0.03$ )	Primary: Hedges' $g=0.64$ (95% CI 0.39 to 0.84, $I^2=0.0\%$ ) Secondary: Hedges' $g=0.88$ (95% CI 0.42 to 1.33, $I^2=0.0\%$ )	The effect size for reduction in depression post-intervention is larger in secondary depression studies, but both subgroups contain little statistical heterogeneity
Change in depression scores	Depression measure: MADRS, QIDS, or BDI	Significant ( $P=0.04$ )	MADRS: Hedges' $g=0.71$ (95% CI $-0.34$ to $1.09$ , $I^2=27.5\%$ ) QIDS: Hedges' $g=0.52$ (95% CI 0.21 to 0.83, $I^2=0.0\%$ ) BDI: Hedges' $g=0.88$ (95% CI 0.42 to 1.33, $I^2=0.0\%$ )	Although studies using the patient reported QIDS and BDI questionnaires showed statistically significant effects for psilocybin, the point estimate for MADRS (clinician assessed) fell between the two The MADRS subgroup displayed a small degree of heterogeneity, whereas the QIDS and BDI subgroups did not
Change in depression scores	Psilocybin dosage: 10-15 mg or 20-25 mg	Not significant ( $P=0.36$ )	10-15 mg: Hedges' $g=0.57$ (95% CI $-0.26$ to $0.87$ , $I^2=0.0\%$ ) 20-25 mg: Hedges' $g=0.73$ (95% CI 0.47 to 1.00, $I^2=0.0\%$ )	The two dosage groups did not appear to differentially affect the primary outcome, change in depression scores. Superior effect sizes seen in the higher dosage group should thus be interpreted with caution
Change in depression scores	Time of assessment post-psilocybin: 2-4 weeks or 4-8 weeks	Not significant ( $P=0.97$ )	2-4 weeks: Hedges' $g=0.67$ (95% CI 0.43 to 0.91, $I^2=0.0\%$ ) 4-8 weeks: Hedges' $g=0.67$ (95% CI 0.25 to 1.08, $I^2=0.0\%$ )	The two time to assessment groups did not appear to differentially affect the primary outcome, change in depression scores
Response rate	Type of depression: primary or secondary	Not significant ( $P=0.09$ )	Primary: risk ratio 1.72 (95% CI 0.81 to 3.68, $I^2=37.0\%$ ) Secondary: risk ratio 3.44 (95% CI 0.05 to 2.50, $I^2=17.6\%$ ), based on two studies only	There are insufficient data to conclude that response rate is affected by type of depression
Remission rate	Type of depression: primary or secondary	Not significant ( $P=0.11$ )	Primary: risk ratio 2.50 (95% CI 1.09 to 5.70, $I^2=7.8\%$ ) Secondary: risk ratio 4.52 (95% CI 0.17 to 11.70, $I^2=0.0\%$ ), based on two studies only	There are insufficient data to conclude that remission rate is affected by type of depression

 $I^2$ =measure of heterogeneity.

BDI=Beck depression inventory; CI=confidence interval; MADRS=Montgomery-Åsberg depression rating scale; QIDS=quick inventory of depressive symptomatology.

robust across timepoints for Ross et al,<sup>17</sup> and were not affected by the splitting of the sample in the case of Goodwin et al<sup>18</sup> and its follow-up,<sup>65</sup> which ensured that multiplicity effects did not affect our conclusions (see supplementary Appendix J).

### Discussion

In our meta-analysis we found that psilocybin use showed a significant benefit on change in depression scores compared with placebo. This is consistent with other recent meta-analyses and trials of psilocybin as a standalone treatment for depression<sup>73 74</sup> or in combination with psychological support.<sup>24 25 29-32 68 75</sup>

This review adds to those finding by exploring treatment effect moderators, with exploratory subgroup analyses showing that the type of depression (primary or secondary) and the depression scale used (Montgomery-Åsberg depression rating scale, Quick Inventory of Depressive Symptomatology, or Beck depression inventory) may have a significant differential effect on the outcome. High between study heterogeneity has been identified by some other meta-analyses of psilocybin (eg, Goldberg et al<sup>29</sup>), with a higher treatment effect in studies with patients with comorbid life threatening conditions compared with patients with primary depression.<sup>22</sup> Although

**Table 5 | Metaregression analyses to explore potential causes of heterogeneity among included studies**

Factor	Outcome	Effect of factor on outcome	Adjusted $R^2$ (%)
Participants with past use of psychedelics (%)	Change in depression scores	Non-significant ( $P=0.38$ )	NA
Participants with past use of psychedelics (%)	Response rate	Significant ( $P=0.001$ )—each percentage increase point increase in past psychedelic use was associated with an increase of 6.0 points in response rates	90.4
Participants with previous use of psychedelics (%)	Remission rate	Significant ( $P=0.024$ )—each percentage increase point increase in past psychedelic use was associated with an increase of 3.9 points in remission rates	98.0
Female participants (%)	Change in depression scores	Non-significant ( $P=0.26$ )	NA
Female participants (%)	Response rate	Non-significant ( $P=0.35$ )	NA
Female participants (%)	Remission rate	Non-significant ( $P=0.25$ )	NA
Average participant's age	Change in depression scores	Non-significant ( $P=0.43$ )	NA
Average participant's age	Response rate	Significant ( $P=0.015$ )—each percentage increase point increase in average age was associated with an increase of 0.13 points in response rates	66.3
Average participant's age	Remission rate	Non-significant ( $P=0.16$ )	NA

NA=not applicable.

possible explanations, including personal factors (eg, patients with life threatening conditions being older) or depression related factors (eg, secondary depression being more severe than primary depression) could be considered, these hypotheses are not supported by baseline data (ie, patients with secondary depression do not differ substantially in age or symptom severity from patients with primary depression). The differential effects from assessment scales used have not been examined in other meta-analyses of psilocybin, but this review's finding that studies using the Beck depression inventory showed a higher treatment effect than those using the Montgomery-Åsberg depression rating scale and Quick Inventory of Depressive symptomatology is consistent with studies in the psychological literature that have shown larger treatment effects when self-report scales are used.<sup>76-77</sup> This finding may be because clinicians tend to overestimate the severity of depression symptoms at baseline assessments, leading to less pronounced differences between before and after treatment identified in clinician assessed scales (eg, Montgomery-Åsberg depression rating scale, Quick Inventory of Depressive symptomatology).<sup>78</sup> However, all subgroup analyses may have been affected by small study biases, and thus the findings should be considered exploratory.

Metaregression analyses further showed that a higher average age and a higher percentage of participants with past use of psychedelics both correlated with a greater improvement in depression scores with psilocybin use and explained a substantial amount of between study variability. For instance, Studerus et al<sup>79</sup> identified participants' age as the only personal variable significantly associated with psilocybin response, with older participants reporting a higher "blissful state" experience. This might be because of older people's increased experience in managing negative emotions and the decrease in 5-hydroxytryptamine type 2A receptor density associated with older age.<sup>80</sup> Furthermore, Rootman et al<sup>81</sup> reported that the cognitive performance of older participants (>55 years) improved significantly more than that of younger participants after micro dosing with psilocybin. Therefore, the higher decrease in depressive symptoms associated with older age could be attributed to a decrease in cognitive difficulties experienced by older participants.

Interestingly, a clear pattern emerged for past use of psychedelics—the higher the proportion of study participants who had used psychedelics in the past, the higher the post-psilocybin treatment effect observed. Past use of psychedelics has been proposed to create an expectancy bias among participants and amplify the positive effects of psilocybin<sup>82-84</sup>; however, this important finding has not been examined in other meta-analyses and may highlight the role of expectancy in psilocybin research.

#### Limitations of this study

Generalisability of the findings of this meta-analysis was limited by the lack of racial and ethnic diversity in the included studies—more than 90% of participants

were white across all included trials, resulting in a homogeneous sample that is not representative of the general population. Moreover, it was not possible to distinguish between subgroups of participants who had never used psilocybin and those who had taken psilocybin more than a year before the start of the trial, as these data were not provided in the included studies. Such a distinction would be important, as the effects of psilocybin on mood may wane within a year after being administered.<sup>21-85</sup> Also, how psychological support was conceptualised was inconsistent within studies of psilocybin interventions; many studies failed to clearly describe the type of psychological support participants received, and others used methods ranging from directive guidance throughout the treatment session to passive encouragement or reassurance (eg, Griffiths et al,<sup>14</sup> Carhart-Harris et al<sup>63</sup>). The included studies also did not gather evidence on participants' previous experiences with treatment approaches, which could influence their response to the trials' intervention. Thus, differences between participant subgroups related to past use of psilocybin or psychotherapy may be substantial and could help interpret this study's findings more accurately. Lastly, the use of graphical extraction software to estimate the findings of studies where exact numerical data were not available (eg, Ross et al<sup>17</sup>), may have affected the robustness of the analyses.

A common limitation in studies of psilocybin is the likelihood of expectancy effects augmenting the treatment effect observed. Although some studies used low dose psychedelics as comparators to deal with this problem (eg, Carhart-Harris et al,<sup>63</sup> Goodwin et al,<sup>18</sup> Griffiths et al<sup>14</sup>) or used a niacin placebo that can induce effects similar to those of psilocybin (eg, Grob et al,<sup>15</sup> Ross et al<sup>17</sup>), the extent to which these methods were effective in blinding participants is not known. Other studies have, however, reported that participants can accurately identify the study groups to which they had been assigned 70-85% of the time,<sup>84-86</sup> indicating a high likelihood of insufficient blinding. This is especially likely for studies in which a high proportion of participants had previously used psilocybin and other hallucinogens, making the identification of the drug's acute effects easier (eg, Griffiths et al,<sup>14</sup> Grob et al,<sup>15</sup> Ross et al<sup>17</sup>). Patients also have expectations related to the outcome of their treatment, expecting psilocybin to improve their symptoms of depression, and these positive expectancies are strong predictors of actual treatment effects.<sup>87-88</sup> Importantly, the effect of outcome expectations on treatment effect is particularly strong when patient reported measures are used as primary outcomes,<sup>89</sup> which was the case in several of the included studies (eg, Griffiths et al,<sup>14</sup> Grob et al,<sup>15</sup> Ross et al<sup>17</sup>). Unfortunately, none of the included studies recorded expectations before treatment, so it is not possible to determine the extent to which this factor affected the findings.

Lastly, a methodological limitation of this review involves the imputation of SD measures from SE values obtained using graphical extraction software when the actual values were not reported in the studies (for the

study of Ross et al<sup>17</sup> in particular). These imputations may lead to inconsistencies between estimates of SE and SD values when different graphical extraction software is used, and thus the analyses should be considered exploratory.

### Implications for clinical practice

Although this review's findings are encouraging for psilocybin's potential as an effective antidepressant, a few areas about its applicability in clinical practice remain unexplored. Firstly, it is unclear whether the protocols for psilocybin interventions in clinical trials can be reliably and safely implemented in clinical practice. In clinical trials, patients receive psilocybin in a non-traditional medical setting, such as a specially designed living room, while they may be listening to curated calming music and are isolated from most external stimuli by wearing eyeshades and external noise-cancelling earphones. A trained therapist closely supervises these sessions, and the patient usually receives one or more preparatory sessions before the treatment commences. Standardising an intervention setting with so many variables is unlikely to be achievable in routine practice, and consensus is considerably lacking on the psychotherapeutic training and accreditations needed for a therapist to deliver such treatment.<sup>90</sup> The combination of these elements makes this a relatively complex and expensive intervention, which could make it challenging to gain approval from regulatory agencies and to gain reimbursement from insurance companies and others. Within publicly funded healthcare systems, the high cost of treatment may make psilocybin treatment inaccessible. The high cost associated with the intervention also increases the risk that unregulated clinics may attempt to cut costs by making alterations to the protocol and the therapeutic process,<sup>91 92</sup> which could have detrimental effects for patients.<sup>92-94</sup> Thus, avoiding the conflation of medical and commercial interests is a primary concern that needs to be dealt with before psilocybin enters mainstream practice.

### Implications for future research

More large scale randomised trials with long follow-up are needed to fully understand psilocybin's treatment potential, and future studies should aim to recruit a more diverse population. Another factor that would make clinical trials more representative of routine practice would be to recruit patients who are currently using or have used commonly prescribed serotonergic antidepressants. Clinical trials tend to exclude such participants because many antidepressants that act on the serotonin system modulate the 5-hydroxytryptamine type 2A receptor that psilocybin primarily acts upon, with prolonged use of tricyclic antidepressants associated with more intense psychedelic experiences and use of monoamine oxidase inhibitors or SSRIs inducing weaker responses to psychedelics.<sup>95-97</sup> Investigating psilocybin in such patients would, however, provide valuable insight on how psilocybin interacts with commonly prescribed

drugs for depression and would help inform clinical practice.

Minimising the influence of expectancy effects is another core problem for future studies. One strategy would be to include expectancy measures and explore the level of expectancy as a covariate in statistical analysis. Researchers should also test the effectiveness of condition masking. Another proposed solution would be to adopt a 2×2 balanced placebo design, where both the drug (psilocybin or placebo) and the instructions given to participants (told they have received psilocybin or told they have received placebo) are crossed.<sup>98</sup> Alternatively, clinical trials could adopt a three arm design that includes both an inactive placebo (eg, saline) and active placebo (eg, niacin, lower psilocybin dose),<sup>98</sup> allowing for the effects of psilocybin to be separated from those of the placebo.

Overall, future studies should explore psilocybin's exact mechanism of treatment effectiveness and outline how its physiological effects, mystical experiences, dosage, treatment setting, psychological support, and relationship with the therapist all interact to produce a synergistic antidepressant effect. Although this may be difficult to achieve using an explanatory randomised trial design, pragmatic clinical trial designs may be better suited to psilocybin research, as their primary objective is to achieve high external validity and generalisability. Such studies may include multiple alternative treatments rather than simply an active and placebo treatment comparison (eg, psilocybin v SSRI v serotonin-noradrenaline reuptake inhibitor), and participants would be recruited from broader clinical populations.<sup>99 100</sup> Although such studies are usually conducted after a drug's launch,<sup>100</sup> earlier use of such designs could help assess the clinical effectiveness of psilocybin more robustly and broaden patient access to a novel type of antidepressant treatment.

### Conclusions

This review's findings on psilocybin's efficacy in reducing symptoms of depression are encouraging for its use in clinical practice as a drug intervention for patients with primary or secondary depression, particularly when combined with psychological support and administered in a supervised clinical environment. However, the highly standardised treatment setting, high cost, and lack of regulatory guidelines and legal safeguards associated with psilocybin treatment need to be dealt with before it can be established in clinical practice.

We thank DT who acted as an independent secondary reviewer during the study selection and data review process.

**Contributors:** AMM contributed to the design and implementation of the research, analysis of the results, and writing of the manuscript. MC was involved in planning and supervising the work and contributed to the writing of the manuscript. AMM and MC are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

**Funding:** None received.

**Competing interests:** All authors have completed the ICMJE uniform disclosure form at <https://www.icmje.org/disclosure-of-interest/> and declare: no support from any organisation for the submitted work;



AMM is employed by IDEA Pharma, which does consultancy work for pharmaceutical companies developing drugs for physical and mental health conditions; MC was the supervisor for AMM's University of Oxford MSc dissertation, which forms the basis for this paper; no other relationships or activities that could appear to have influenced the submitted work.

**Ethical approval:** This study was approved by the ethics committee of the University of Oxford Nuffield Department of Medicine, which waived the need for ethical approval and the need to obtain consent for the collection, analysis, and publication of the retrospectively obtained anonymised data for this non-interventional study.

**Data sharing:** The relevant aggregated data and statistical code will be made available on reasonable request to the corresponding author.

**Transparency:** The corresponding author (AMM) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as registered have been explained.

**Dissemination to participants and related patient and public communities:** To disseminate our findings and increase the impact of our research, we plan on writing several social media posts and blog posts outlining the main conclusions of our paper. These will include blog posts on the websites of the University of Oxford's Department of Primary Care Health Sciences and Department for Continuing Education, as well as print publications, which are likely to reach a wider audience. Furthermore, we plan to present our findings and discuss them with the public in local mental health related events and conferences, which are routinely attended by patient groups and advocacy organisations.

**Provenance and peer review:** Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- World Health Organization. Depressive Disorder (Depression); 2023. <https://www.who.int/news-room/fact-sheets/detail/depression>.
- James SL, Abate D, Abate KH, et al, GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018;392:1789-858. doi:10.1016/S0140-6736(18)32279-7
- Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391:1357-66. doi:10.1016/S0140-6736(17)32802-7
- 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. doi:10.1176/ajp.2006.163.11.1905
- Mitchell AJ. Two-week delay in onset of action of antidepressants: new evidence. *Br J Psychiatry* 2006;188:105-6. doi:10.1192/bjp.bp.105.011692
- Bockting CL, Hollon SD, Jarrett RB, Kuyken W, Dobson K. A lifetime approach to major depressive disorder: The contributions of psychological interventions in preventing relapse and recurrence. *Clin Psychol Rev* 2015;41:16-26. doi:10.1016/j.cpr.2015.02.003
- Nierenberg AA, Petersen TJ, Alpert JE. Prevention of relapse and recurrence in depression: the role of long-term pharmacotherapy and psychotherapy. *J Clin Psychiatry* 2003;64(Suppl 15):13-7.
- Ling S, Ceban F, Lui LMW, et al. Molecular mechanisms of psilocybin and implications for the treatment of depression. *CNS Drugs* 2022;36:17-30. doi:10.1007/s40263-021-00877-y
- Tyš F, Páleníček T, Horáček J. Psilocybin--summary of knowledge and new perspectives. *Eur Neuropsychopharmacol* 2014;24:342-56. doi:10.1016/j.euroneuro.2013.12.006
- Carbonaro TM, Bradstreet MP, Barrett FS, et al. Survey study of challenging experiences after ingesting psilocybin mushrooms: Acute and enduring positive and negative consequences. *J Psychopharmacol* 2016;30:1268-78. doi:10.1177/0269881116662634
- Carhart-Harris RL, Bolstridge M, Rucker J, et al. Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *Lancet Psychiatry* 2016;3:619-27. doi:10.1016/S2215-0366(16)30065-7
- Carhart-Harris RL, Bolstridge M, Day CMJ, et al. Psilocybin with psychological support for treatment-resistant depression: six-month follow-up. *Psychopharmacology (Berl)* 2018;235:399-408. doi:10.1007/s00213-017-4771-x
- Davis AK, Barrett FS, May DG, et al. Effects of psilocybin-assisted therapy on major depressive disorder: a randomized clinical trial. *JAMA Psychiatry* 2021;78:481-9. doi:10.1001/jamapsychiatry.2020.3285
- 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. doi:10.1177/0269881116675513
- Grob CS, Danforth AL, Chopra GS, et al. Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. *Arch Gen Psychiatry* 2011;68:71-8. doi:10.1001/archgenpsychiatry.2010.116
- Kraehenmann R, Preller KH, Scheidegger M, et al. Psilocybin-induced decrease in amygdala reactivity correlates with enhanced positive mood in healthy volunteers. *Biol Psychiatry* 2015;78:572-81. doi:10.1016/j.biopsych.2014.04.010
- 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. doi:10.1177/0269881116675512
- Goodwin GM, Aaronson ST, Alvarez O, et al. Single-dose psilocybin for a treatment-resistant episode of major depression. *N Engl J Med* 2022;387:1637-48. doi:10.1056/NEJMoa2206443
- Bogenschutz MP, Johnson MW. Classic hallucinogens in the treatment of addictions. *Prog Neuropsychopharmacol Biol Psychiatry* 2016;64:250-8. doi:10.1016/j.pnpbp.2015.03.002
- Bogenschutz MP, Podrebarac SK, Duane JH, et al. Clinical interpretations of patient experience in a trial of psilocybin-assisted psychotherapy for alcohol use disorder. *Front Pharmacol* 2018;9:100. doi:10.3389/fphar.2018.00100
- Carhart-Harris RL, Roseman L, Bolstridge M, et al. Psilocybin for treatment-resistant depression: fMRI-measured brain mechanisms. *Sci Rep* 2017;7:13187. doi:10.1038/s41598-017-13282-7
- Galvão-Coelho NL, Marx W, Gonzalez M, et al. Classic serotonergic psychedelics for mood and depressive symptoms: a meta-analysis of mood disorder patients and healthy participants. *Psychopharmacology (Berl)* 2021;238:341-54. doi:10.1007/s00213-020-05719-1
- Dos Santos RG, Osório FL, Crippa JA, Riba J, Zuardi AW, Hallak JE. Antidepressive, anxiolytic, and antiaddictive effects of ayahuasca, psilocybin and lysergic acid diethylamide (LSD): a systematic review of clinical trials published in the last 25 years. *Ther Adv Psychopharmacol* 2016;6:193-213. doi:10.1177/2045125316638008
- Ko K, Kopra EJ, Cleare AJ, Rucker JJ. Psychedelic therapy for depressive symptoms: A systematic review and meta-analysis. *J Affect Disord* 2023;322:194-204. doi:10.1016/j.jad.2022.09.168
- Romeo B, Karila L, Martelli C, Benyamina A. Efficacy of psychedelic treatments on depressive symptoms: A meta-analysis. *J Psychopharmacol* 2020;34:1079-85. doi:10.1177/0269881120919957
- Vollenweider FX, Preller KH. Psychedelic drugs: neurobiology and potential for treatment of psychiatric disorders. *Nat Rev Neurosci* 2020;21:611-24. doi:10.1038/s41583-020-0367-2
- Roseman L, Demetriou L, Wall MB, Nutt DJ, Carhart-Harris RL. Increased amygdala responses to emotional faces after psilocybin for treatment-resistant depression. *Neuropharmacology* 2018;142:263-9. doi:10.1016/j.neuropharm.2017.12.041
- Daws RE, Timmermann C, Gribaldi B, et al. Increased global integration in the brain after psilocybin therapy for depression. *Nat Med* 2022;28:844-51. doi:10.1038/s41591-022-01744-z
- Goldberg SB, Pace BT, Nicholas CR, Raison CL, Hutson PR. The experimental effects of psilocybin on symptoms of anxiety and depression: A meta-analysis. *Psychiatry Res* 2020;284:112749. doi:10.1016/j.psychres.2020.112749
- Irizarry R, Winczura A, Dimassi O, Dhillon N, Minhas A, Larice J. Psilocybin as a Treatment for Psychiatric Illness: A Meta-Analysis. *Cureus* 2022;14:e31796. doi:10.7759/cureus.31796
- Johnson MW, Griffiths RR. Potential therapeutic effects of psilocybin. *Neurotherapeutics* 2017;14:734-40. doi:10.1007/s13311-017-0542-y
- Castro Santos H, Gama Marques J. What is the clinical evidence on psilocybin for the treatment of psychiatric disorders? A systematic review. *Porto Biomed J* 2021;6:e128. doi:10.1097/pbj.0000000000000128
- Li NX, Hu YR, Chen WN, Zhang B. Dose effect of psilocybin on primary and secondary depression: a preliminary systematic review and meta-analysis. *J Affect Disord* 2022;296:26-34. doi:10.1016/j.jad.2021.09.041

- 34 Yu CL, Liang CS, Yang FC, et al. Trajectory of antidepressant effects after single- or two-dose administration of psilocybin: A systematic review and multivariate meta-analysis. *J Clin Med* 2022;11:938. doi:10.3390/jcm11040938
- 35 Moreno FA, Wiegand CB, Taitano EK, Delgado PL. Safety, tolerability, and efficacy of psilocybin in 9 patients with obsessive-compulsive disorder. *J Clin Psychiatry* 2006;67:1735-40.
- 36 Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264-9, W64. doi:10.7326/0003-4819-151-4-200908180-00135
- 37 Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. doi:10.1136/bmj.l4898
- 38 Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *J Clin Epidemiol* 2011;64:380-2. doi:10.1016/j.jclinepi.2010.09.011
- 39 Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002. doi:10.1136/bmj.d4002
- 40 Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. John Wiley & Sons, 2019. doi:10.1002/9781119536604
- 41 Andrade C. Mean difference, standardized mean difference (SMD), and their use in meta-analysis: as simple as it gets. *J Clin Psychiatry* 2020;81:11349. doi:10.4088/JCP.20f13681
- 42 Lin L, Aloe AM. Evaluation of various estimators for standardized mean difference in meta-analysis. *Stat Med* 2021;40:403-26. doi:10.1002/sim.8781
- 43 Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 2010;1:97-111. doi:10.1002/jrsm.12
- 44 DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* 2007;28:105-14. doi:10.1016/j.cct.2006.04.004
- 45 Borenstein M, Hedges L, Rothstein H. Meta-analysis: Fixed effect vs. random effects. *com*. 2007;1-62.
- 46 Int'Hout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* 2016;6:e010247. doi:10.1136/bmjopen-2015-010247
- 47 Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. *Stat Med* 2007;26:1964-81. doi:10.1002/sim.2688
- 48 Tendal B, Nüesch E, Higgins JP, Jüni P, Gøtzsche PC. Multiplicity of data in trial reports and the reliability of meta-analyses: empirical study. *BMJ* 2011;343:d4829. doi:10.1136/bmj.d4829
- 49 Spinelli LM, Pandis N. Prediction interval in random-effects meta-analysis. *Am J Orthod Dentofacial Orthop* 2020;157:586-8. doi:10.1016/j.ajodo.2019.12.011
- 50 Higgins JP, Green S. Identifying and measuring heterogeneity. *Cochrane handbook for systematic reviews of interventions*. 2011;5(0).
- 51 Austin PC, Brunner LJ. Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. *Stat Med* 2004;23:1159-78. doi:10.1002/sim.1687
- 52 O'Donnell KC, Mennenga SE, Bogenschütz MP. Psilocybin for depression: Considerations for clinical trial design. *J Psychedelic Stud* 2019;3:269-79. doi:10.1556/2054.2019.026.
- 53 Baker WL, Cios DA, Sander SD, Coleman CI. Meta-analysis to assess the quality of warfarin control in atrial fibrillation patients in the United States. *J Manag Care Pharm* 2009;15:244-52. doi:10.18553/jmcp.2009.15.3.244
- 54 Berlin JA, Santanna J, Schmid CH, Szczec LA, Feldman HI, Anti-Lymphocyte Antibody Induction Therapy Study Group. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002;21:371-87. doi:10.1002/sim.1023
- 55 Iyengar S, Greenhouse J. Sensitivity analysis and diagnostics. *Handbook of research synthesis and meta-analysis*. Russell Sage Foundation, 2009:417-33.
- 56 Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int J Surg* 2021;88:105906. doi:10.1016/j.ijsu.2021.105906
- 57 Griffiths R, Barrett F, Johnson M, Mary C, Patrick F, Alan D. Psilocybin-Assisted Treatment of Major Depressive Disorder: Results From a Randomized Trial. Proceedings of the ACNP 58th Annual Meeting: Poster Session II. In *Neuropsychopharmacology*. 2019;44:230-384.
- 58 Barrett F. ACNP 58th Annual Meeting: Panels, Mini-Panels and Study Groups. [Abstract.] *Neuropsychopharmacology* 2019;44:1-77. doi:10.1038/s41386-019-0544-z.
- 59 Benville J, Agin-Lieb G, Roberts DE, et al. Effects of psilocybin on suicidal ideation in patients with life-threatening cancer. *Biol Psychiatry* 2021;89:S235-6. doi:10.1016/j.biopsych.2021.02.591.
- 60 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. doi:10.1177/02698811211073759
- 61 Griffiths RR, Hurwitz ES, Davis AK, Johnson MW, Jesse R. Survey of subjective "God encounter experiences": Comparisons among naturally occurring experiences and those occasioned by the classic psychedelics psilocybin, LSD, ayahuasca, or DMT. *PLoS One* 2019;14:e0214377. doi:10.1371/journal.pone.0214377
- 62 Doss MK, Považan M, Rosenberg MD, et al. Psilocybin therapy increases cognitive and neural flexibility in patients with major depressive disorder. *Transl Psychiatry* 2021;11:574. doi:10.1038/s41398-021-01706-y
- 63 Carhart-Harris R, Giribaldi B, Watts R, et al. Trial of psilocybin versus escitalopram for depression. *N Engl J Med* 2021;384:1402-11. doi:10.1056/NEJMoa2032994
- 64 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. doi:10.1192/bjo.2022.565
- 65 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. doi:10.1016/j.jad.2023.01.108
- 66 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.
- 67 Ross S, Agin-Lieb G, Lo S, et al. Acute and sustained reductions in loss of meaning and suicidal ideation following psilocybin-assisted psychotherapy for psychiatric and existential distress in life-threatening cancer. *ACS Pharmacol Transl Sci* 2021;4:553-62. doi:10.1021/acspsci.1c00020
- 68 Vargas AS, Luís Á, Barroso M, Gallardo E, Pereira L. Psilocybin as a new approach to treat depression and anxiety in the context of life-threatening diseases—a systematic review and meta-analysis of clinical trials. *Biomedicine* 2020;8:331. doi:10.3390/biomedicine8090331
- 69 Sterne JA, Egger M. Regression methods to detect publication and other bias in meta-analysis. In: *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments*. John Wiley and Sons, 2005;99-110.
- 70 Isojarvi J, Wood H, Lefebvre C, Glanville J. Challenges of identifying unpublished data from clinical trials: Getting the best out of clinical trials registers and other novel sources. *Res Synth Methods* 2018;9:561-78. doi:10.1002/jrsm.1294
- 71 Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002. doi:10.1136/bmj.d4002
- 72 Clarke M, Brice A, Chalmers I. Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLoS One* 2014;9:e102670. doi:10.1371/journal.pone.0102670
- 73 Hodge AT, Sukpraput-Braaten S, Narlesky M, Strahan RC. The use of psilocybin in the treatment of psychiatric disorders with attention to relative safety profile: a systematic review. *J Psychoactive Drugs* 2023;55:40-50. doi:10.1080/02791072.2022.2044096
- 74 Prouzeau D, Conejero I, Voyvodic PL, Becamel C, Abbar M, Lopez-Castromar J. Psilocybin efficacy and mechanisms of action in major depressive disorder: a review. *Curr Psychiatry Rep* 2022;24:573-81. doi:10.1007/s11920-022-01361-0
- 75 Więckiewicz G, Stoklosa I, Piega M, Górczka P, Pudło R. Lysergic acid diethylamide, psilocybin and dimethyltryptamine in depression treatment: a systematic review. *Pharmaceuticals (Basel)* 2021;14:793. doi:10.3390/ph14080793
- 76 John Mann J, Ellis SP, Currier D, et al. Self-rated depression severity relative to clinician-rated depression severity: trait stability and potential role in familial transmission of suicidal behavior. *Arch Suicide Res* 2016;20:412-25. doi:10.1080/13811118.2015.1033504
- 77 Zimmerman M, Walsh E, Friedman M, Boerescu DA, Attiullah N. Are self-report scales as effective as clinician rating scales in measuring treatment response in routine clinical practice? *J Affect Disord* 2018;225:449-52. doi:10.1016/j.jad.2017.08.024
- 78 Dorz S, Borgherini G, Conforti D, Scarso C, Magni G. Comparison of self-rated and clinician-rated measures of depressive symptoms: a naturalistic study. *Psychol Psychother* 2004;77:353-61. doi:10.1348/1476083041839349
- 79 Studerus E, Gamma A, Komter M, Vollenweider FX. Prediction of psilocybin response in healthy volunteers. *PLoS One* 2012;7:e30800. doi:10.1371/journal.pone.0030800
- 80 Adams KH, Pinborg LH, Svarer C, et al. A database of [(18)F]-altanserin binding to (2A) receptors in normal volunteers: normative data and relationship to physiological and demographic variables. *Neuroimage* 2004;21:1105-13. doi:10.1016/j.neuroimage.2003.10.046

- 81 Rootman JM, Kiraga M, Kryskow P, et al. Psilocybin microdoses demonstrate greater observed improvements in mood and mental health at one month relative to non-microdosing controls. *Sci Rep* 2022;12:11091. doi:10.1038/s41598-022-14512-3
- 82 Zeiss R, Gahr M, Graf H. Rediscovering psilocybin as an antidepressive treatment strategy. *Pharmaceuticals (Basel)* 2021;14:985. doi:10.3390/ph14100985
- 83 Turner EH, Rosenthal R. Efficacy of antidepressants. *BMJ* 2008;336:516-7. doi:10.1136/bmj.39510.531597.80
- 84 Bershad AK, Schepers ST, Bremmer MP, Lee R, de Wit H. Acute subjective and behavioral effects of microdoses of lysergic acid diethylamide in healthy human volunteers. *Biol Psychiatry* 2019;86:792-800. doi:10.1016/j.biopsych.2019.05.019
- 85 Barrett FS, Doss MK, Sepeda ND, Pekar JJ, Griffiths RR. Emotions and brain function are altered up to one month after a single high dose of psilocybin. *Sci Rep* 2020;10:2214.
- 86 Carbonaro TM, Johnson MW, Hurwitz E, Griffiths RR. Double-blind comparison of the two hallucinogens psilocybin and dextromethorphan: similarities and differences in subjective experiences. *Psychopharmacology (Berl)* 2018;235:521-34. doi:10.1007/s00213-017-4769-4
- 87 Dupuis D. Psychedelics as tools for belief transmission. set, setting, suggestibility, and persuasion in the ritual use of hallucinogens. *Front Psychol* 2021;12:730031. doi:10.3389/fpsyg.2021.730031
- 88 Horvath AO, Del Re AC, Flückiger C, Symonds D. Alliance in individual psychotherapy. *Psychotherapy (Chic)* 2011;48:9-16. doi:10.1037/a0022186
- 89 Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry* 2013;170:723-33. doi:10.1176/appi.ajp.2012.12040474
- 90 Pearson C, Siegel J, Gold JA. Psilocybin-assisted psychotherapy for depression: Emerging research on a psychedelic compound with a rich history. *J Neurol Sci* 2022;434:120096. doi:10.1016/j.jns.2021.120096
- 91 Harding L. Regulating Ketamine Use in Psychiatry. *J Am Acad Psychiatry Law* 2023;51:320-5.
- 92 Zhang MW, Hong YX, Husain SF, Harris KM, Ho RC. Analysis of print news media framing of ketamine treatment in the United States and Canada from 2000 to 2015. *PLoS One* 2017;12:e0173202. doi:10.1371/journal.pone.0173202
- 93 George JR, Michaels TI, Sevelius J, Williams MT. The psychedelic renaissance and the limitations of a White-dominant medical framework: A call for indigenous and ethnic minority inclusion. *J Psychedelic Stud* 2020;4:4-15doi:10.1556/2054.2019.015.
- 94 Michaels TI, Purdon J, Collins A, Williams MT. Inclusion of people of color in psychedelic-assisted psychotherapy: a review of the literature. *BMC Psychiatry* 2018;18:245. doi:10.1186/s12888-018-1824-6
- 95 Bonson KR, Buckholtz JW, Murphy DL. Chronic administration of serotonergic antidepressants attenuates the subjective effects of LSD in humans. *Neuropsychopharmacology* 1996;14:425-36. doi:10.1016/0893-133X(95)00145-4
- 96 Yamauchi M, Miyara T, Matsushima T, Imanishi T. Desensitization of 2A receptor function by chronic administration of selective serotonin reuptake inhibitors. *Brain Res* 2006;1067:164-9. doi:10.1016/j.brainres.2005.10.075
- 97 Coleshill MJ, Sharpe L, Colloca L, Zachariae R, Colagiuri B. Placebo and active treatment additivity in placebo analgesia: research to date and future directions. *Int Rev Neurobiol* 2018;139:407-41. doi:10.1016/bs.irn.2018.07.021
- 98 Aday JS, Heifets BD, Pratscher SD, Bradley E, Rosen R, Woolley JD. Great Expectations: recommendations for improving the methodological rigor of psychedelic clinical trials. *Psychopharmacology (Berl)* 2022;239:1989-2010. doi:10.1007/s00213-022-06123-7
- 99 Sugarman J, Califf RM. Ethics and regulatory complexities for pragmatic clinical trials. *JAMA* 2014;311:2381-2. doi:10.1001/jama.2014.4164
- 100 Sox HC, Lewis RJ. Pragmatic trials: practical answers to "real world" questions. *JAMA* 2016;316:1205-6. doi:10.1001/jama.2016.11409

#### Supplementary information: Appendices A-J