Dear Editors,

We thank you for your time and greatly appreciate your feedback on our manuscript. Please see our responses to each of the Committee's and Reviewers' comments below:

Comments by the Committee:

COMMENT: We were wondering whether you would be interested in making this a living systematic review.

RESPONSE: Given that the field of psychedelic research is fast-moving and new research is being published at a high rate, we would be interested in making this a living systematic review and discussing how this might be accommodated by the BMJ.

COMMENT: It's a very dense and crowded paper and the paper needs to be made much more readable.

RESPONSE: We have revised the manuscript to remove some text and moved content to the Supplemental Materials section (e.g., Reasons for Exclusion table). However, we have tried to retain as much information as possible, because of the comments of the reviewers about the manuscript being well-written, comprehensive, and easy to follow. However, we would welcome guidance from the editors on further changes that could increase the manuscript's readability; for example, to remove some of our detailed and comprehensive descriptions of the included studies.

COMMENT: We were concerned that the gaps in the literature were not clearly identified and that the findings are similar to other systematic reviews.

RESPONSE: We have outlined and clarified the added value of our research in the "Gaps in research on psilocybin" section on page 5 of the manuscript. We have further noted that the strengths of this review include a larger sample size and the consideration of covariates that were not assessed in any other reviews. These elements have led to the elucidation of new findings not outlined in other reviews, such as that prior use of psychedelics may have a significant effect on depression score changes, as highlighted by the first reviewer. We would welcome further suggestions if these points should be emphasised more.

COMMENT: It's not clear what Hedge's g is, something more transparent would be preferable

RESPONSE: Hedge's g is the most commonly used measure of effect size in other meta-analyses of psilocybin and psychedelic research (Galvão-Coelho et al., 2021; Goldberg et al., 2020; Irizarry et al., 2022; Li et al., 2022; Yu et al., 2022). We prefer this measure because Hedge's g accounts for small sample bias (which Cohen's d does not account for) (Borenstein et al., 2009), and many of the included studies have small samples (<20 participants). We have added a note that a Hedge's g value of 0.2, 0.5, 0.8 or 1.2 corresponds to a small, medium, large, and very large effect respectively (Hedges and Olkin, 1985), to help readers interpret these values within the context of the manuscript's findings on page 23.

COMMENT: Should use baseline-adjusted change in outcome rather than simple change.

RESPONSE: The baseline-adjusted change would be meaningful if the results are likely to change substantially based on the baseline value, but this was not supported by the data in this or other systematic reviews (e.g., Galvão-Coelho et al., 2021; Irizarry et al., 2022; Kiseley et al., 2022; Ko et al., 2022; Li et al., 2022; Yu et al., 2022). Further, given that only group average change values were available in the included studies, making this calculation would not be possible without collecting more data from the original researchers.

COMMENT: Please define direction of rating scales.
RESPONSE: We have added clarification on the direction of the rating scales on page 8.

COMMENT: Methods appear in Results.
RESPONSE: We have moved the relevant sections into the Methods.

COMMENT: It was not clear what the finding was really- so you say, "The meta-analysis showed a significant benefit for psilocybin (Hedge's g=1.64, 95%CI: 0.55 to 2.73, p<0.001)..." and then you say "...in favor of placebo" for the main outcome. Please clarify.

RESPONSE: We have amended this section on page 2.

Comments from Reviewers

Reviewer: 1

Comments:
The manuscript is very well written.
The literature reviewed is updated, and as far as I know, the trial has been conducted according to the standards for this kind of analysis.
This method and results parts are clear, however some part of the manuscript might need to be slightly revised.
Please find below some comments for different parts of the manuscript.

Abstract- Clear
Keywords- Clear
COI: clear
Intro: well written, informative.

COMMENT 1- Introduction p.4
Current evidence on psilocybin's effectiveness for depressive symptoms.
Authors do not really discuss the dose of psilocybin (10.1016/j.euroneuro.2023.07.011), It would be valuable to consider incorporating information related to the optimal dosage and its relevance to specific patient subgroups. This could be included in either the introduction or the discussion section of the paper to provide a more comprehensive understanding of the study's implications

RESPONSE: We have added a consideration about the optimal effective dosage on page 38. We have not expanded further on this topic in order to avoid adding to the complexity of the manuscript, as suggested by the committee.

Methodology- very clear

COMMENT 2-
Was the study's methodology inclusive of control groups utilizing psychotherapy, and were antidepressants considered within the control condition, which could encompass non-active comparators like placebo, niacin, or psychedelic microdosing, as mentioned?

RESPONSE: As outlined on pages 6-7, studies were eligible for inclusion if psychotherapy was given to both the control and active groups for the study. Studies that used antidepressants for the control group were also eligible, because antidepressants are considered to be the standard of care for depression and are not classed as psychedelic medication.

COMMENT 3- Authors should be aware that psilocybin can have an effect, starting at doses as low as 3mg/70kg
'' Studies where the active psilocybin condition involves micro dosing (i.e., psilocybin below 100 µg/kg) were excluded.''

RESPONSE: We are aware of different estimates of the dose threshold at which psilocybin has shown effects and chose a commonly accepted threshold, similar to the reviews of Galvão-Coelho et al., 2021 and Prouzeau et al., 2022, and other trials of psilocybin (e.g., Cavanna et al., 2022; van Elk et al., 2021).

COMMENT 4- Search Strategy
Were articles screened independently by two different authors?

RESPONSE: Articles were screened independently by AM and another student (DT), whose contribution is highlighted in the Acknowledgements section.

COMMENT 5
‘’Subgroup analysis’’ ‘’ Number of doses and amount of psilocybin administered’’
The authors should exercise caution in distinguishing between a psychedelic-naïve population and participants who may have had prior experiences with psilocybin more than a year before being enrolled in trials. It's crucial to acknowledge that the antidepressant effects of psilocybin are reported to endure for at least six months but tend to wane after a year. This differentiation and recognition of the temporal aspects are crucial for a more accurate interpretation of the study's outcomes.

RESPONSE: We have made a note about the difference between a psychedelic naïve population and participants who may have had prior experiences with psilocybin more than a year before being enrolled in trials in the Discussion (page 34). However, data on these two populations were not provided in the included studies, so an analysis of separate groups was not possible.

COMMENT 6
‘’ Reasons for exclusion’’> I suggest that this part could go to the supplements

RESPONSE: We have moved this section to the supplements (Appendix F)

COMMENT 7
Table: How do authors define psychological support?
The authors should consider a comprehensive discussion of this point in the paper's discussion section. It is challenging to assess the potential influence of prior experiences with psychotherapy, psychological support, or active shamanic sessions on the patients included in the study. Exploring this aspect in the discussion would provide valuable insights into the nuanced factors that may affect the outcomes and interpretations of the research.

RESPONSE: We have made a note about the different ways that psychological support is defined in psychedelic research in the Discussion (page 34). However, we have not expanded on this point to avoid adding to the complexity of the manuscript, as suggested by the committee.

COMMENT 8
‘’ Side effects and adverse events observed across studies’’
The is one trial where a patient had vivid dreams and insomnia lasting more than 24 hours and required medication and maybe one case of some thought disturbance if I remember. Check (10.1016/j.euroneuro.2023.07.011). Bipolar/schizophrenia patients might be more susceptible to such adverse events.
More largely side-effects are reported with different scales and methods in each trials…

RESPONSE: We have added a note about this side-effect on page 20 and Appendix E. We have not expanded on side-effects in the manuscript given how different scales and methods are used to report them across trials. However, we have added a note about this on page 20.

COMMENT 9- ‘’ The percentage of participants with prior psychedelics use may have a significant effect on depression score changes (p=0.002),’’ This is an important finding that should be highlighted.

RESPONSE: We have highlighted this finding in the Abstract and page 32. We welcome suggestions from the editors on how to highlight this point further, if needed.

Reviewer: 2

Comments:
I think this study has several strengths. It was well written and easy to follow. The risk of bias assessment and GRADE evaluation were helpful as was the evaluation of response and remission rates. It was preregistered through PROSPERO. It addresses a timely topic. These strengths aside, I have some questions about the study and its reporting that I think are worth considering. I hope the authors and editor find these helpful.

COMMENT 1. The primary question / concern that arose for me is whether this study adds information that moves substantially beyond other meta-analyses that have been published on this topic, including a recent meta-analysis focusing specifically on psilocybin for depression (Haikazian et al., 2023, Psychiatry Research). Indeed, it appears that the Haikazian et al. meta-analysis actually includes a trial that I believe would be eligible for inclusion in the current meta-analysis but was not included I'm guessing because it appeared after the current study's search was completed (Raison et al., 2023, JAMA). The current meta-analysis provides GRADE assessment and tests some moderators not explored in the Haikazian et al. paper. Whether the current meta-analysis is sufficiently novel to warrant publication in BMJ is, of course, an editorial decision.

RESPONSE: Our review differs from Haikazian et al.'s (2023) meta-analysis in the following ways:
• We searched a wider variety of international databases and trial registries, and searched for both published and unpublished data. Haikazian et al. (2023) limited their search to published English-language literature only (searching PubMed, Embase, PsycINFO, and MedLine).
• We evaluated each individual study using the Cochrane risk-of-bias tool for randomized trials (RoB 2), as well as the totality of evidence for each outcome using the GRADE assessment approach. Haikazian et al. (2023) only used the RoB 2 tool and did not appraise the totality of data for each outcome.
• Haikazian et al. (2023) included the study of Davis et al. (2021), which we excluded because psychotherapeutic support was only provided to patients in the psilocybin treatment group, making it impossible to separate the effects of psilocybin from the effects of psychotherapy on depression.
• We examined a range of moderating factors, including type of depression (primary or secondary), prior experience with psychedelics, psilocybin dosage, type of outcome measure (clinician-rated or self-reported), and demographics (e.g., age, gender). We identified significant moderating relationships for a few of these factors and we discussed recommendations for further research. Haikazian et al. (2023) only considered a subgroup analysis for patients with primary and secondary depression.
As the reviewer suggests, Raison et al.'s (2023) study was not included in the manuscript because it was published after the completion of our search – including it post hoc without conducting and processing a complete new search might lead to bias. However, we have considered it in the discussion section (page 31) and would also consider it as a candidate for inclusion if this becomes a living systematic review.

COMMENT 2. I appreciated the authors preregistering their study in PROSPERO. However, it appeared to me that the preregistration as reported in supplemental materials was slightly different from that on the PROSPERO website. For example, the subgroup analysis in the manuscript is not entirely consistent with the "analysis of subgroups or subsets" section on the PROSPERO record.

RESPONSE:The differences between the subgroup analyses outlined in the PROSPERO registration and the manuscript are minor. For example, the PROSPERO record notes that subgroup analyses will be conducted on high and low risk of bias studies. However, because we did not identify any 'high-risk' studies, such an analysis was not possible. As noted in the analysis of subgroups or subsets section, "These subgroup analyses will be conducted given that a sufficient number of studies are identified within each subgroup", so any deviations from the planned analyses are due to a lack of relevant data.

COMMENT 3. I was curious how the authors determined whether there were sufficient studies to conduct meta-regression. I may have missed it, but I did not see this specified in the preregistration.

RESPONSE: We used the common convention of 10 studies being sufficient (based on the Cochrane Handbook's recommendations), as mentioned on page 11 of the manuscript.

COMMENT 4. I was curious if the authors managed to obtain results from studies that were completed but not yet published. It would be helpful to more clearly report how many such studies existed.

RESPONSE: As outlined on pages 12-13 ("Requesting more information" section) we reached out to authors to inquire about data presented in conferences. No other unpublished studies eligible for inclusion were identified during our search of grey literature.

COMMENT 5. It was not clear to me why the authors reviewed some secondary outcomes (e.g., White Bear Suppression Inventory) that were not depression specific.

RESPONSE: We have removed discussion of such outcomes.

COMMENT 6. I think it is reasonable to include the studies that looked at depression in the context of terminal cancer (although I think one could also argue that end-of-life depression is distinct from typical depression / MDD). However, I wondered to what extent tests of moderation were confounded with there being much larger effects for these trials (e.g., d = 4.52 in Griffiths et al., 2016). For example, were the end-of-life samples also older?

RESPONSE: We have highlighted the point that studies of secondary depression showed considerably higher effect sizes for reduction in depression post-intervention than studies of primary depression on page 31. Although there is no evidence that this is caused by a difference in average age, as participants with terminal conditions tend to be the same age as patients with primary depression, we have added a note in the Discussion about the need to explore this finding further (page 31). We have not expanded on this, to avoid adding to the complexity of the manuscript, as suggested by the committee.