Appendices

Appendix A: PROSPERO registered study protocol

"The efficacy of psilocybin for the treatment of depressive symptoms; a systematic review and meta-analysis"

Rationale and background

Depression

Depression affects an estimated 300 million people around the world, with that number having increased by nearly 20% over the last decade (World Health Organization, 2019). Depression is also the number one cause of disability worldwide (James et al., 2018), and increases the relative risk of all-cause mortality twofold (Walker et al., 2015) while the economic burden of depression is estimated to be $200 billion per year in the US alone (Greenberg et al., 2015).

Currently, pharmacotherapies for depression are widely available; the most prominent class of drugs, selective serotonin reuptake inhibitors, (SSRIs), act by increasing levels of brain monoamine neurotransmitters such as serotonin and norepinephrine through monoamine reuptake inhibition (Finkel et al., 2008). However, these pharmacotherapies appear to have limited efficacy, can have serious adverse effects, and are associated with low patient adherence (Cipriani et al. 2018; Rush et al., 2006). It is estimated that 30-50% of people with depression gain no benefit from pharmacotherapies (Hengartner and Plöderl 2018; Munkholm et al. 2019). Additionally, 10-30% of patients are classed as 'treatment-resistant', meaning that they have not responded to at least two distinct pharmacological interventions (Akil et al. 2018; Posternak and Zimmerman 2005). Importantly, the therapeutic effects of anti-depressant pharmacotherapies only occur 5-7 weeks after the start of the treatment course, and it often takes months for remission of symptoms to be achieved (Rush et al., 2006; Carvalho et al. 2016). Thus, it is necessary to explore new treatments for depression with mechanisms of action different to those of classical SSRIs, to both improve treatment

response rate and decrease the time it takes to see improvements in depressive symptoms following administration.

Psilocybin

A substance that has been explored as a potential pharmacological treatment for depression is Psilocybin (4-phosphoryloxy-*N*, *N*-dimethyltryptamine), a naturally occurring serotonergic hallucinogen found in several species of mushrooms throughout the world (Tylš et al. 2014; McKenna & Riba, 2016). Psilocybin was first isolated from the *P. Mexicana* mushroom by Swiss chemist Albert Hofmann in 1958 and has since then been synthetically produced (McKenna and Riba, 2016).

Psilocybin, along with other psychedelic substances, is being investigated as a treatment for depression because it is an agonist of serotonin 5-$HT_{1A/2A/2C}$ receptors (Vollenweider and Kometer, 2010). The activation of fronto-cortical glutamate receptors secondary to 5-HT_{2A} receptor-related glutamate release appears to be the key mechanism of action, while intravenous administration of psilocybin significantly decreases blood flow to the medial prefrontal cortex and the posterior cingulate cortex; the same normalisation of hyperactivity is observed with classical anti-depressant treatments (Dos Santos et al., 2016).

The psychedelic effects of psilocybin, which usually take effect at doses over 20 mg, involve an altered consciousness, with increased introspection and hypnagogic experiences (Patra, 2016). Perceptual changes such as synaesthesia, delusions, and alterations in the sense of time are also observed (Dos Santos et al., 2016). Importantly, significant and long-term improvements in wellbeing, positive mood, and optimism have also been reported after such psychedelic experiences (Griffiths et al., 2008; Hendricks et al., 2015).

Despite promising early studies, clinical research with psilocybin was interrupted in the late 1960s, as the recreational use of hallucinogenic substances and their association with countercultural movements prompted the marginalization and defunding of psychedelic research across the US and Europe (Hintzen and Passie, 2010; Oram, 2014). Thus, while psilocybin's therapeutic potential has been known for a long time, it only recently that rigorous clinical research has started being conducted again.

Effects of psilocybin on depressive symptoms

The renewed interest in psilocybin's anti-depressive effects has led to the conduction of several clinical trials on treatment-resistant depression (Carhart-Harris et al., 2016, 2018), major depressive disorder (Davis et al., 2021), and physical illness-related depression (Griffiths et al., 2016; Grob et al., 2011; Kraehenmann et al. 2015; Ross et al., 2016). These trials have mostly reported positive efficacy findings, showing reductions in depressive symptoms after the administration of psilocybin. Critically, only minimal adverse effects have been reported in these studies, and drug harm assessments of healthy volunteers indicate that psilocybin does not induce physiological toxicity, is not addictive, and does not lead to withdrawal (Bogenschutz and Johnson 2016; Bogenschutz and Ross 2018).

Why is it important to do this review?

Several systematic reviews and meta-analyses have investigated the effectiveness of psilocybin as a treatment for depressive symptoms within the last decade, most of which found encouraging results for psilocybin's treatment potential. However, some of these reviews included both healthy volunteers and depressive patients (Galvão-Coelho et al., 2021), and most have combined multiple serotonergic psychedelics (Romeo et al., 2020; Wieckiewicz et al., 2021), which may not be methodologically appropriate (Yu et al., 2022). Additionally, a few systematic reviews have included non-randomised studies and studies in which psilocybin was administered in conjunction with psychotherapeutic interventions (Goldberg et al., 2020; Wieckiewicz et al., 2021). Lastly, no systematic reviews conducted so far have considered grey literature (Li et al., 20222, Yu et al., 2022), which might have led to a significant overestimation of psilocybin's therapeutic effects.
Thus, this dissertation will be a systematic review of both indexed and non-indexed RCTs exploring the effectiveness of psilocybin in people with depressive symptoms.

Objective

The aim of this dissertation is to complete a systematic review and meta-analysis of indexed and non-indexed randomised controlled trials (RCTs) investigating the effectiveness of psilocybin administration in treating depressive symptoms, in comparison with a placebo or non-psychoactive pharmacological agent.

Population = people with depressive symptoms as measured by validated clinician or subject-rated scales
Intervention = psilocybin administration (excluding microdosing)
Comparison = placebo / non-psychoactive pharmacological agent
Outcome = changes in depressive symptoms as measured by validated clinician or subject-rated scales

Inclusion Criteria

1. **Study type**
   RCTs with a cross-over or parallel design will be eligible for inclusion. Only studies with a control condition will be included; this can be any type of non-active comparator, such as placebo, niacin, or psychedelic micro dosing. Observational studies, systematic reviews, meta-analyses, qualitative studies, opinion pieces, letters, editorials, case reports, and animal studies will be excluded.
2. **Participants**
   The study's sample should be comprised of people with depressive symptoms, evaluated using a clinically validated tool for depression and mood disorder outcomes. Such tools include the Beck Depression Inventory (BDI), Hamilton Depression Rating Scale (HAM-D), Montgomery–Åsberg Depression Rating Scale (MADRS), Positive and Negative Affect Schedule (PANAS), Profile of Mood States (POMS), and Quick Inventory of Depressive Symptomatology (QIDS). Studies of participants who experience depressive symptoms comorbid with physical conditions (e.g., cancer) will also be included. However, studies of healthy participants (without depressive symptomatology) will not be eligible for inclusion.
3. **Intervention**
   Eligible studies will investigate the effect of psilocybin on depressive symptoms. Studies where the active psilocybin condition involves directive psychological therapy, psilocybin-assisted psychotherapy, or micro dosing (i.e., psilocybin below 100 μg/kg) will also be excluded.
4. **Comparator**
   Any inactive comparator is eligible for inclusion, including placebo, niacin, and micro doses of psychedelics.
5. **Outcome**

Changes in depressive symptoms, measured by validated scales, such as those listed in the "Participants" section will be considered. Outcomes will be included irrespective of the time point at which the measurements were taken, as the effects of psilocybin start within hours of administration. However, studies that measured the outcome less than 3 hours after the administration of psilocybin will be excluded, as any reported changes could be attributed to the transient cognitive and affective effects of administering the substance.

Search strategy

The major electronic databases and trial registries of psychological and medical research will be searched, with no limits placed on the publication date (e.g., Clinical Trials.gov, Cochrane Library, Embase, Medline, ProQuest, Scopus, Web of Science, PsychInfo etc.). Searching through multiple databases is necessary, as there is no complete overlap between them, and each database includes journals not indexed in the others.

Unpublished literature should be searched through registries of past and ongoing trials, databases of conference proceedings, government reports, theses, dissertations, and grant registries (e.g., ClinicalTrials.gov, PsycEXTRA).

While no language restrictions will be placed on online searches, a dedicated search of **non-English databases** will also be conducted (e.g., DRKS, LILACS), with the help of a translator.

Reference and bibliography lists of eligible studies will also be hand searched for relevant publications.

A search update will be conducted if the period between running the searches and the analysis of the data exceeds six months.

Data collection, extraction, and management

Literature search results will be imported to the Endnote X9 reference management software, and following the removal of duplicate results, the references will be imported to the Covidence platform. There, each reference's title and abstract will be screened by the primary

and secondary reviewers, who will then also screen the full text of the remaining references. Any disagreement about eligibility will be resolved through discussion among the reviewers. If there is insufficient information to determine eligibility, the study's authors will be contacted. The reviewers will not be blinded to the studies' authors, institutions, or the journal in which they are published.

The study selection process and reasons for excluding studies that were considered for full-text eligibility will be presented in a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta- Analyses) flow diagram (Moher et al., 2009)

The Stata 16 software will be used if a meta-analysis is indicated.

The Systematic Review Data Repository (SRDR http://srdr.ahrq.gov/), produced by the Agency for Healthcare Research and Quality (AHRQ) will be used to create a tailored, computerized, online data extraction form using PRISMA guidelines.

Critical appraisal of individual studies

The methodological quality of the eligible studies will be explored using the Cochrane Collaboration's Risk of Bias 2 tool (RoB2) for assessing risk of bias in randomised trials (Higgins et al., 2011c), which includes an assessment of Selection bias, Performance bias, Detection bias, Attrition bias, and Reporting bias. Each criterion will be given a 'low risk', 'high risk, or 'unclear risk' rating, and the results will be presented in a 'Risk of bias' table. No summary score will be extracted.
In addition to the criteria specified by the RoB2 Tool, the potential impact of industry funding and conflicts of interest on reported outcomes will also be considered.

Critical appraisal of the aggregated evidence

The overall methodological quality of the aggregated evidence will be evaluated using the validated and robust Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines. These guidelines allow for the appraisal of the evidence in terms of factors not included in the RoB tool, such as inconsistency, imprecision, indirectness, confounding, effect magnitude and publication bias (Guyatt et al., 2011), providing a more holistic assessment of the evidence's quality.

Further, if a meta-analysis is conducted and evidence of heterogeneity among the trials is found, publication bias will be assessed using a funnel plot and asymmetry tests (e.g., Egger's test) (Sterne et al., 2011).

Data synthesis

The extracted data will be summarised in a series of tables outlining the demographic, methodological and outcome-related characteristics of the included studies. A narrative synthesis of the characteristics and results of the included studies will also be provided. A meta-analysis will also be performed if >5 studies meet the inclusion criteria and have sufficient homogeneity. The meta-analysis will be carried out with the Stata 16 software.

Investigation of heterogeneity and sensitivity analysis

Statistical heterogeneity will be tested using the Chi-squared test (significance level: 0.1) and I^2 statistic, and heterogeneity among reviewed studies will also be evaluated visually and displayed graphically using a Forest plot.
If substantial or considerable levels of heterogeneity are found (I^2 >=50% or P <0.1) (Higgins and Green, 2011) the study design and characteristics of the included studies will be analysed. The source of heterogeneity will be explored by subgroup analysis, and its potential effects on the results discussed.
A **sensitivity analysis** will be performed if there are any studies considered to be at high risk of bias and to investigate the effect of the inclusion of any unpublished studies. Exclusion sensitivity plots will be used to graphically display the impact of individual studies, and to determine which studies have a particularly large influence on the results of the meta-analysis. All sensitivity analyses will be carried out with the Stata 16 software.

Subgroup analysis

To reduce the risk of errors caused by multiplicity and to avoid data-fishing, subgroup analyses will be planned *a priori* and will be limited to:

- Patient characteristics, including age and sex, as the remain mostly unexplored in published research.

- Comorbidities, such as a serious physical condition. Prior research indicates that the effects of psilocybin may be less strong for such participants, compared to participants with no comorbid physical conditions (Li et al., 2022).
- Number of doses and amount of psilocybin administered, as some previous meta-analyses found that a higher dose level and a higher number of doses of psilocybin both predicted a greater reduction in depressive symptoms (Yu et al., 2022), while others reported the opposite pattern (Li et al., 2022).
- Clinician vs subject-rated scales
- High vs low quality studies
- If heterogeneity in the pooled effect estimates is very low, subgroup analyses may not be performed.

Meta-regression

If there are sufficient studies, meta-regression will be carried out to investigate whether covariates, or potential effect modifiers, explain any of the heterogeneity of effects between studies. Meta-regression will be carried out with the Stata 16 software.

Note: the choice of a random or fixed effects model for the meta-analysis will be partly decided based on the degree of heterogeneity found among the studies (Tufanaru et al., 2015), and thus is not pre-specified in the protocol

Appendix B: Search syntax

1 Depression/

2 mood disorders/ or exp depressive disorder/

3 depress*.mp. [mp=title, book title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

4 1 or 2 or 3

5 Psilocybin/

6 psilocyb*.mp. [mp=title, book title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

7 "psychedelic*".m_titl.

8 5 or 6 or 7

9 4 and 8

10 randomized controlled trial.pt.

11 controlled clinical trial.pt.

12 randomized.ab.

13 drug therapy.fs.

14 randomly.ab.

15 trial.ab.

16 groups.ab.

17 10 or 11 or 12 or 13 or 14 or 15 or 16

18 exp animals/ not humans.sh.

19 17 not 18

20 9 and 19

Appendix C: Data extraction template

Extraction template – adapted from the Cochrane Data collection form for intervention reviews (RCTs)

General Information

Reviewer Name	
Extraction Date	
Study ID and Report ID	
Authors	
Study Source	
Study Type (e.g., full paper, conference abstract etc.)	
Notes	

Study Eligibility

Study Characteristics	Eligibility criteria	Eligibility criteria met? Yes	No	Unclear	Not reported	Location in text or source *(pg & ¶/fig/table/other)* and relevant text
Type of study	Randomised Controlled Trial (open label or double-blind)	☐	☐	☐		
Participants	People with depressive symptoms (evaluated with validated tool)	☐	☐	☐		
Types of intervention	Psilocybin administration	☐	☐	☐		
Types of comparison	Inactive comparator (placebo, niacin, and micro doses of psychedelics)	☐	☐	☐		
Types of outcome measures	Changes in depressive symptoms, measured by validated scales, at least >3 hours post-administration of psilocybin	☐	☐	☐		

INCLUDE ☐	EXCLUDE ☐
Reason for exclusion	
Notes:	

PROCEED ONLY IF STUDY MEETS ELIGIBILITY CRITERIA

Characteristics of included studies

	Description as stated in report/paper	Location in text or source
Country of correspondence author		
Country where study conducted		
Language of publication		
Study design		
Funding source		
Dates study began and ended		
No of centers		
Ethical approval obtained	YES/NO	
Notes:		

Participants

	Description as stated in report/paper	Location in text or source
Population description (e.g., diagnostic status)		
Setting		
Inclusion criteria		
Exclusion criteria		

Method of recruitment of participants		
Informed consent obtained	☐ Yes ☐ No ☐ Unclear	
Total no. assessed for eligibility		
Total no. recruited		
Total no. randomised		
Total no. of withdrawals and exclusions		
Reasons for withdrawal or exclusion		
Clusters *(no., type, no. people per cluster)*		
Baseline imbalances		
Age (mean, SD, range)		
Sex (proportion)		
Race/ethnicity (proportion)		
Severity of depressive illness		
Co-morbidities		
Other relevant sociodemographics		
Subgroups measured		
Subgroups reported		
Notes:		

Intervention Groups

	Description as stated in report/paper	Location in text or source
Group name		
No. randomised to group		

Demographics of group (Age, Race/ethnicity, Sex)		
Description *(e.g. content, dose, components)*		
Duration of treatment period		
Timing *(e.g. frequency, duration of each episode)*		
Delivery *(e.g. mechanism, medium, intensity, fidelity)*		
Providers *(e.g. no., profession, training, ethnicity etc. if relevant)*		
Co-interventions		
Integrity of delivery		
Compliance		
Notes:		

Outcomes (for each outcome, including side-effects)

	Description as stated in report/paper	Location in text or source
Outcome name		
Measurement tool(s)		
Time points measured		
Time points reported		
Outcome definition		
Person measuring/ reporting		
Unit of measurement		
Scales: upper and lower limits *(indicate whether high or low score is good)*		

Is outcome/tool validated?	☐ Yes ☐ No ☐ Unclear		
Imputation of missing data *(e.g. assumptions made for ITT analysis)*			
Power *(e.g. power & sample size calculation, level of power achieved)*			
Notes:			

Data extraction (for each outcome)

	Description as stated in report/paper						Location in text or source
Outcome							
Measuring tool							
Subgroup							
Time point(s)							
Post-intervention or change from baseline?							
Results	Intervention			Comparison			
	Mean	SD *(or other variance, specify)*	No. participants	Mean	SD *(or other variance, specify)*	No. participants	
Any other results reported *(e.g. mean difference, CI, P value)*							
No. missing participants							
Reasons missing							

Statistical methods used and appropriateness of these *(e.g. adjustment for correlation)*				
Reanalysis required? *(specify)*	☐ Yes	☐ No	☐ Unclear	
Reanalysis possible?	☐ Yes	☐ No	☐ Unclear	
Reanalysed results				
Notes:				

Other information

	Description as stated in report/paper	Location in text or source
Key conclusions of study authors		
References to other relevant studies		
Study funding sources		
Possible conflicts of interest		
Correspondence information		
Notes, including anything that needs to requested from the authors:		

Appendix D: Upcoming trials eligible for inclusion in the review

Title	Registration Code	Expected Completion Date	Trial registry
Psilocybin for Treatment-Resistant Depression	NCT05029466	February 2023	ClinicalTrials.gov
Efficacy and Safety of Psilocybin in Treatment-Resistant Major Depression (EPIsoDE)	NCT04670081	March 1, 2024	ClinicalTrials.gov
Psilocybin Treatment of Major Depressive Disorder With Co-occurring Alcohol Use Disorder (PsiloMDDAUD)	NCT04620759	August 31, 2026	ClinicalTrials.gov
Psilocybin in Depression Resistant to Standard Treatments (PsiDeR)	NCT04959253	November 1, 2023	ClinicalTrials.gov
Efficacy, Safety, and Tolerability of a Single Administration of COMP360 in Participants With TRD	NCT05624268	October 2024	ClinicalTrials.gov
The Effect of Psilocybin on MDD Symptom Severity and Synaptic Density (PSIPET)	NCT04630964	February 6, 2024	ClinicalTrials.gov
PAPR: PAP + MBSR for Front-line Healthcare Provider COVID-19 Related Burnout (PAPR)	NCT05557643	January 1, 2024	ClinicalTrials.gov
Frontline Clinician Psilocybin Study	NCT05163496	March 30, 2024	ClinicalTrials.gov

Psilocybin for Major Depressive Disorder	NCT05675800	March 31, 2025	ClinicalTrials.gov
Psilocybin - Induced Neuroplasticity in the Treatment of Major Depressive Disorder	NCT03554174	April 2024	ClinicalTrials.gov
Long Term Follow Up Study to COMP 001 And COMP 003 Trials (P-TRD LTFU)	NCT04519957	Unknown	ClinicalTrials.gov
Psilocybin-assisted psychotherapy for the treatment of depression and anxiety associated with life-threatening illness	ACTRN12619001225101	Post May 2023	WHO ICTRP
Psilocybin versus ketamine – fast acting antidepressant strategies in treatment-resistant depression	2018-004480-31	Unknown	WHO ICTRP
Psilocybin - a strategy of rapid antidepressant response in depression comorbid with cancer, a randomized double-blind study with the possibility of entering open extension.	2020-005037-32	Unknown	WHO ICTRP

Appendix E: Reasons for exclusion from the systematic review

Study	Reason for exclusion
Agin-Liebes et al. (2020) (45) (follow-up of Ross et al., 2016 (17))	Participants were followed up after the initial treatment crossover, so they had all received both psilocybin and placebo (niacin) and were analysed as a single group
Carhart-Harris et al. (2016) (11) and its follow-up Carhart-Harris et al. (2018) (12)	no control group, all participants received active psilocybin treatment
Shnayder et al. (2023) (46)	no control group, all participants received active psilocybin treatment
Roseman et al. (2018) (47)	Neuroimaging data, all participants received active psilocybin treatment with no control group
Daws et al. (2022) (48)	Only neuroimaging data reported, based on Carhart-Harris et al.'s (2021) sample
Doss et al. (2021**)** (44)	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
Davis et al. (2021) (13) and its follow-up Gukasyan et al. (2022) (42)	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
Barba et al. (2022) (49)	Excluded from the meta-analysis, because primary outcomes (change in rumination and thought suppression) reflect specific symptoms of depression, rather than overall depressive mood
Ross et al. (2021) (50)	Excluded from the meta-analysis, because primary outcomes (change in demoralisation) reflect specific symptoms of depression, rather than overall depressive mood

Appendix F: Side effects and adverse events reported in the included studies

Study	Side effects and adverse events reported
von Rotz et al. (2023)	• No events of symptomatic hypertension, extreme anxiety, suicidal behaviour, or psychotic/delusional decompensation were documented, and no rescue medication was used during the trial period of approximately one month • All adverse events (clinically relevant symptoms outlasting acute drug effects) recorded were reported to be mild; 4 cases of headache and the 2 cases of dizziness were categorised as "likely related" due to the temporal relationship to the intervention, 1 case of diarrhoea was already prevalent one day before drug administration but slightly intensified and continued thereafter and was thus categorised as "probably related", and 2 cases of common cold and 1 case of cystitis in the placebo group were classified as "unlikely to be related" to the intervention. • Overall, psilocybin was found to have a mild adverse event profile; the most frequently reported adverse events was mild headache (11%) which resolved completely within two days after drug administration. The incidence of headache in this study was substantially lower than the rate for mild to moderate headache (24–60%) reported in other clinical trials in MDD using higher doses of psilocybin
Goodwin et al. (2022), Goodwin et al. (2023)	• Adverse events occurred in 66 participants (84%) in the 25-mg group, 56 (75%) in the 10-mg group, and 57 (72%) in the 1-mg group. The most frequent adverse events reported in the 25-mg group with onset on the day of psilocybin administration (day 1) were headache (in 24% of the participants), nausea (in 22%), and dizziness and fatigue (in 6% each). Adverse events that were rated as severe on day 1 were reported by 4% of the participants in the 25-mg group, 8% of those in the 10-mg group, and 1% of those in the 1-mg group. Just one participant (in the 25-mg group) was treated with adjunctive medication (lorazepam

	for acute anxiety) on day 1. There were no serious adverse events reported on day 1. • From day 2 up to week 3 (primary endpoint assessment timepoint), severe adverse events were reported by 9% of the participants in the 25-mg group, 7% of those in the 10-mg group, and 1% of those in the 1-mg group. The serious adverse events in the 25-mg group were suicidal ideation (in two participants) and intentional self-injury (nonsuicidal self-injurious behaviour) (in two participants) and in the 10-mg group were suicidal ideation (in two participants), intentional self- injury (in one participant), and hospitalization (for severe depression, in one participant). No serious adverse events were reported from day 2 up to week 3 in the 1-mg group. • After week 3 and up to week 12 (end of trial), severe adverse events were reported by 3% of the participants in the 25-mg group, 4% of those in the 10-mg group, and no participants in the 1-mg group. Serious adverse events in the 25-mg group were suicidal behaviour (in three participants), codeine withdrawal syndrome (in one participant), and adjustment disorder with anxiety and depressed mood (in one participant); in the 10-mg group were intentional self-injury (in one participant), depression (in one participant), and suicidal ideation (in one participant); and in the 1-mg group were intentional self-injury (in one participant). • The number of participants who showed worsening of suicidal state from baseline to week 3 were 11 (14%) in the 25-mg group, 13 (17%) in the 10-mg group, and 7 (9%) in the 1-mg group. 3 participants in the 25-mg group reported suicidal behaviour after week 3. All three had a history of suicidal behaviour or nonsuicidal self-injury before the trial and did not have a treatment response at week 3. No clinically significant changes in vital signs, clinical laboratory tests, or 12-lead ECGs were observed during the trial
Carhart-Harris et al. (2021)	• No serious adverse events were observed in this study. The percentage of patients reporting adverse events was similar in the two groups: 26 (87%) in the psilocybin group and 24 (83%) in the escitalopram group.

	The percentage of patients who had increased anxiety and dry mouth was higher in the escitalopram group than in the psilocybin group. Adverse events in the psilocybin group typically occurred within 24 hours after the dosing day; the most common adverse event was headache. One case of abnormal dreams, illusion and insomnia was reported in the psilocybin group, at an equivalent rate to the escitalopram group. • When cued to report on specific emotional and side-effect–related phenomena through the PTCS, patients in the psilocybin group reported greater perceived improvements in the ability to cry and feel compassion, intense emotion, and pleasure and reported feeling less drowsy than those in the escitalopram group. No cases of visual perceptual changes, psychotic symptoms, or dependency-related behaviours were observed or reported in either trial group at 6 weeks.
Grob et al. (2011)	• No adverse physiological or psychological effects attributable to the treatment were observed. All participants tolerated the treatment sessions well, with no indication of severe anxiety or a "bad trip." The fact that psilocybin produced only modest effects on the anxious ego dissolution scale of the 5D-ASC confirmed this conclusion.
Griffiths et al. (2016)	• No serious adverse events attributed to psilocybin administration occurred. • Consistent with previous research (Griffiths et al., 2006, 2011), there were transient moderate increases in systolic and/or diastolic blood pressure after psilocybin administration. An episode of elevated systolic blood pressure (>160 mm Hg at one or more time-point) occurred in 34% of participants in the high-dose session and 17% of participants in the low-dose session. An episode of elevated diastolic blood pressure (>100 mm Hg at one or more time-point) occurred in 13% of participants in the high-dose session and 2% of participants in the low- dose session. None of these episodes met criteria for medical intervention. • Nausea or vomiting occurred in 15% of participants in the high-dose session and none in the low-dose session. An episode of physical

	discomfort (any type) occurred in 21% of participants in the high-dose session and 8% in the low-dose session. • Also consistent with previous research (Griffiths et al., 2006, 2011), transient episodes of psychological distress during psilocybin sessions were more common after the high dose than the low dose. Psychological discomfort (any type) occurred in 32% of participants in the high-dose session and 12% in the low-dose session. An episode of anxiety occurred in 26% of participants in the high-dose session and 15% in the low-dose session. One participant had a transient episode of paranoid ideation (2% of high-dose sessions). There were no cases of hallucinogen persisting perception disorder or prolonged psychosis. One participant reported mild headache starting toward the end of the high-dose session and lasting until 9 p.m. that evening. Of the 11 participants for whom headache was assessed on the day after sessions, two reported a delayed moderate headache after the high-dose session.
Ross et al. (2016)	• No serious adverse events (medical or psychiatric) attributed to either psilocybin or niacin. No pharmacological interventions (e.g., benzodiazepines, anti-psychotics) were needed during dosing sessions, no participants abused or became addicted to psilocybin, there were no cases of prolonged psychosis or hallucinogen persisting perceptual disorder (HPPD), and no participants required psychiatric hospitalization. • In terms of adverse events attributable to psilocybin, the most common medical adverse events were non-clinically significant elevations in BP and HR (76%), headaches/migraines (28%), and nausea (14%). The most common psychiatric adverse events were transient anxiety (17%) and transient psychotic-like symptoms (7%: one case of transient paranoid ideation and one case of transient thought disorder). These medical and psychiatric adverse events are all known adverse events of psilocybin, were transient, tolerable, and consistent with prior trials of psilocybin administration in normal volunteers (Griffiths et al., 2006, 2008, 2011), and patients with terminal cancer (Grob et al., 2011).

Appendix G: Details of the Risk of Bias Assessment of the included studies

Risk of bias for Primary Depression

All studies provided detailed descriptions of appropriate randomization methods, either in the primary paper or supplementary materials, and no study reported baseline differences indicative of problems with this process. Although participants were likely aware of their assigned intervention, particularly when psilocybin was compared to placebo or waitlist, no deviations from the intended intervention due to the trial context were reported. Outcome data were available for a very high proportion of participants across studies and outcomes related to depressive symptoms and mood were measured using appropriate and clinically validated scales (whether clinician-rated or patient-rated). Reported results were consistent with those pre-specified in the studies' protocols. Finally, all studies were either funded by pharmaceutical companies or led by researchers who served as board members or received consulting fees from biopharmaceutical companies.

Risk of bias for Secondary Depression

Most crossover studies, except Grob et al. (2011) (15), provided detailed descriptions of the randomization methods used, and no baseline differences were reported. As only pre-crossover data was used for this review, studies were not scored for the 'Risk of carryover effects' domain but some concerns about the possibility of bias in relation to carryover effects were noted and clinically significant effects on depressive symptoms and mood have been observed months following the administration of single psilocybin doses, so it is unclear whether the 5-7 week washout period used in the included studies would be sufficient to eliminate carryover effects before participants start the second period of such trials. Participants and investigators were aware of (or were able to guess) the assigned intervention during each period of Grob et al.'s (2011) (15) and Ross et al.'s (2016) (17) trials, but no deviations from the intended intervention arose because of trial context. However, concerns were noted for Griffiths et al. (2016) (14) given that the active psilocybin dose was reduced from 30 to 22mg/70kg and the inactive psilocybin dose from 3 to 1mg/70kg in light of data from dose-finding studies published while the trial was ongoing, but a plan to re-consider dosing during the trial is not mentioned in the paper or supplementary materials. For all

studies reported results were consistent with those pre-specified in the studies' protocols. And outcome data were available for a very high proportion of participants across studies. Outcomes related to depressive symptoms and mood were measured using appropriate and clinically validated scales (whether clinician-rated or patient-rated). Finally, all studies were funded by privately funded non-profit research institutions.

Appendix H: Exploring causes of heterogeneity

Subgroup analyses – Continuous data

The overall between-groups heterogeneity when studies were grouped based on the type of depression investigated (primary or secondary) was significant (p=0.002), indicating that depression type may impact the primary outcome (change in depression scores). Specifically, although both primary and secondary depression studies significantly favour the psilocybin intervention, the effect size for reduction in depression post-intervention is considerably larger in secondary depression studies (g=3.25, 95%CI: 0.97 to 5.53) than in primary depression studies (g=0.84, 95%CI: 0.07 to 1.61) (Figure 5). Nevertheless, both subgroups contain considerable statistical heterogeneity (I^2=79.9% for primary and I^2=79.1% for secondary depression), suggesting that factors other than depression type contribute to the observed heterogeneity.

The overall between-groups heterogeneity when studies were grouped based on the depression measure used (MADRS, QIDS or BDI) was significant (p=0.001), indicating that the choice of depression scale may impact the primary outcome (change in depression scores). Specifically, although studies using the patient reported QIDS and BDI questionnaires significantly favour the psilocybin intervention, the reduction in depression post-intervention is larger in BDI studies (g=3.25, 95%CI: 0.97 to 5.53) than in QIDS studies (g=0.53, 95%CI: 0.03 to 1.02) (Figure 6). In contrast, for studies using the clinician-assessed MADRS questionnaire, reduction in depression scores was not significant following psilocybin treatment (g=1.18, 95%CI: -1.36 to 3.73). Importantly, while both the MADRS and BDI subgroups display considerable heterogeneity (I^2=89.6% and I^2=79.1%, respectively), no significant heterogeneity was observed in the QIDS group (I^2=0.0%, p>0.05).

The overall between-groups heterogeneity when studies were grouped based on the psilocybin dosage administered (10-15mg or 20-25mg) was not significant (p=0.257), indicating that dosage does not appear to impact the primary outcome (change in depression scores). Both subgroups were characterised by considerable heterogeneity (I^2=86.9% for the 10-15mg group and I^2=92.2% for the 20-25mg group) and, although studies using a moderate psilocybin dose (10-15mg) did not show a significant reduction in depression scores (g=1.10, 95%CI: -0.43 to 2.62), unlike those using a higher psilocybin dose (20-25mg) (g=2.10,

95%CI: 0.18 to 4.02), this should be interpreted with caution given the lack of a statistically significant difference between these two effect estimates (Figure 7).

Lastly, the overall between-groups heterogeneity when studies were grouped based on the time of assessment following psilocybin administration (2-4 weeks or 4-8 weeks) was not significant (p=0.279), indicating that time to assessment did not appear to impact the primary outcome (change in depression scores). Both subgroups were characterised by considerable heterogeneity (I^2=82.1% for the 2-4 weeks group and I^2=96.1% for the 4-8 weeks group) and, although studies assessing depression 2-4 weeks following psilocybin administration showed a significant reduction in depression scores (g=1.21, 95% CI: 0.19 to 2.24), unlike those assessing depression later (4-8 weeks post-administration) (g=2.62, 95%CI: -2.66 to 7.90), this should be interpreted with caution given the lack of a statistically significant difference between these two effect estimates (Figure 8).

Subgroup analyses – Dichotomous data

Response

The overall Response rate is shown in Figure 9. Due to the smaller number of studies included in this part of the analysis, heterogeneity was further explored only in terms of the type of depression investigated in each study (primary or secondary). Between-groups heterogeneity appeared to be non-significant (p=0.236), indicating that the likelihood of treatment response was not substantially impacted by type of depression (Figure 10).

Remission

The overall Response rate is shown in Figure 11. The smaller number of studies included in this part of the analysis meant that heterogeneity was further explored for type of depression only (primary or secondary). Between-groups heterogeneity was non-significant (p=0.923), indicating that the likelihood of remission was not substantially impacted by type of depression (Figure 12).

Meta regression

The percentage of female participants did not have a statistically significant effect on the observed change in depression scores (p=0.41), remission (p=0.78) or response (p=0.71).

The average age of participants appeared to have a significant effect on depression score changes ($p<0.001$), with each year increase in age associated with a 0.16 increase in depression score change post-intervention, with the actual increase ranging from 0.08 to 0.24 points. The adjusted R-square indicates that participants' average age can explain 48.5% of between-study variability, providing a possible explanation for the high heterogeneity observed among studies. However, participants' average age did not have a significant association with remission ($p=0.78$) or response ($p=0.12$) post psilocybin administration.

The percentage of participants with prior psychedelics use may have a significant effect on depression score changes ($p=0.002$), with each percentage point increase associated with an increase of 4.2 points in depression score change post-intervention, with the actual increase ranging from 1.5 to 6.9 points. The adjusted R-square indicates that prior psychedelics use can explain 38.9% of between-study variability in depression scores, providing a possible explanation for the high heterogeneity observed among studies. Further, the association between prior psychedelics use and response likelihood was significant ($p=0.006$), with each percentage increase being associated with an increase of 4.4 points in response rates, with the actual increase ranging from 1.3 to 7.5 points. The adjusted R-square indicates that between-study variability in between-studies response likelihood can be almost entirely explained by this factor ($R^2_{adj}=93.8\%$). Prior psychedelics use did not have a significant association with remission ($p=0.15$) post psilocybin administration.

Figure 5. Forest plot for the pre- to post-treatment change in depression scores by depression type

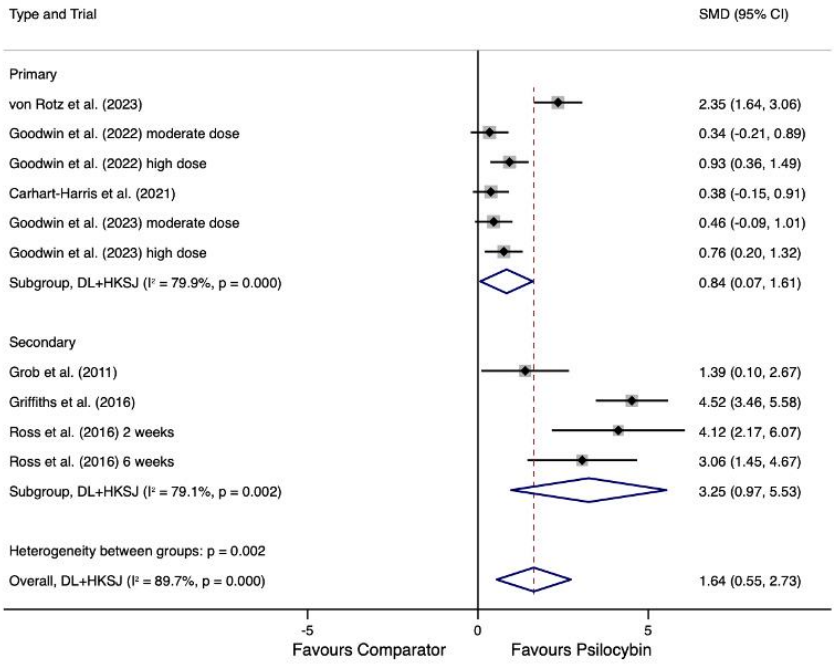


Figure 6. Forest plot for the pre- to post-treatment change in depression scores by symptom measurement scale

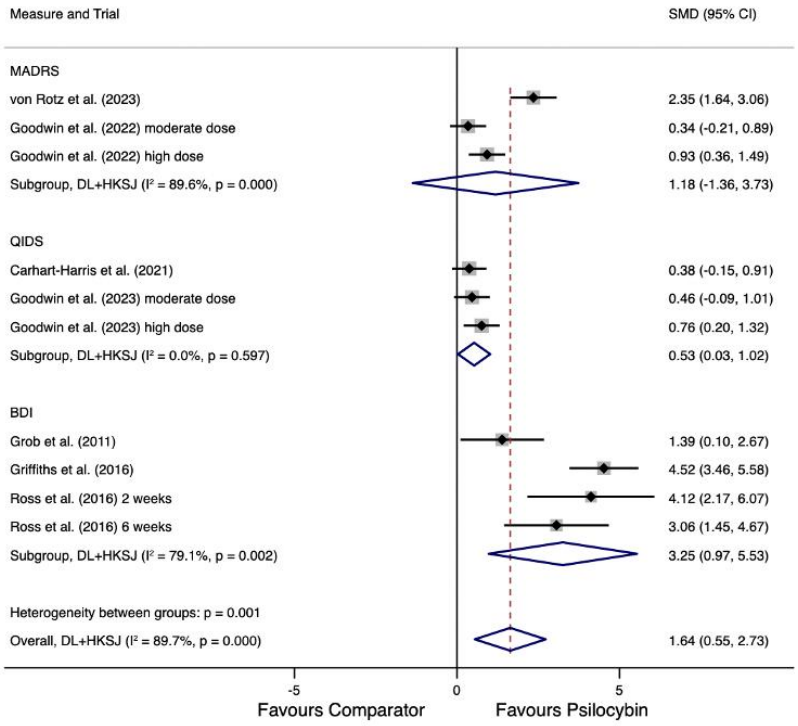


Figure 7. Forest plot for the pre- to post-treatment change in depression scores by psilocybin dosage

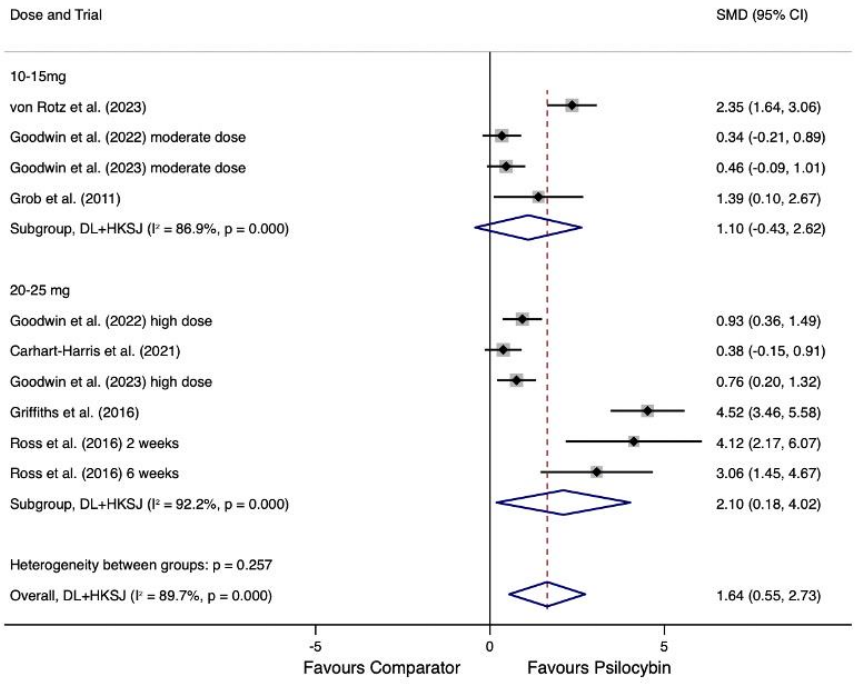


Figure 8. Forest plot for the pre- to post-treatment change in depression scores by time of assessment

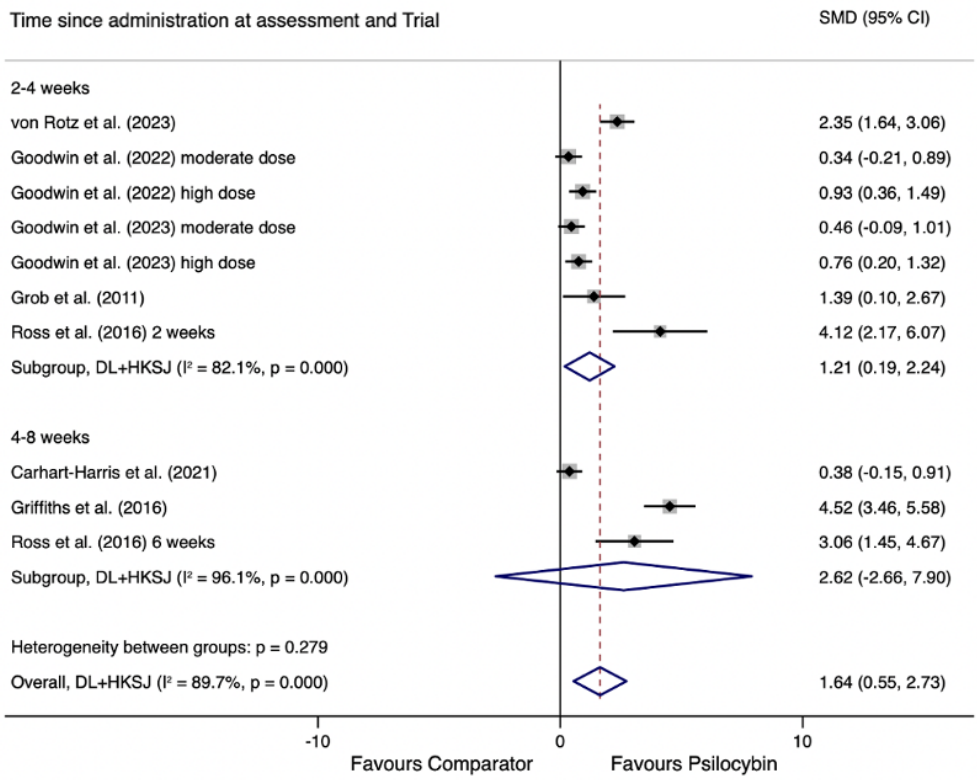


Figure 9. Forest plot for the overall response rate to treatment

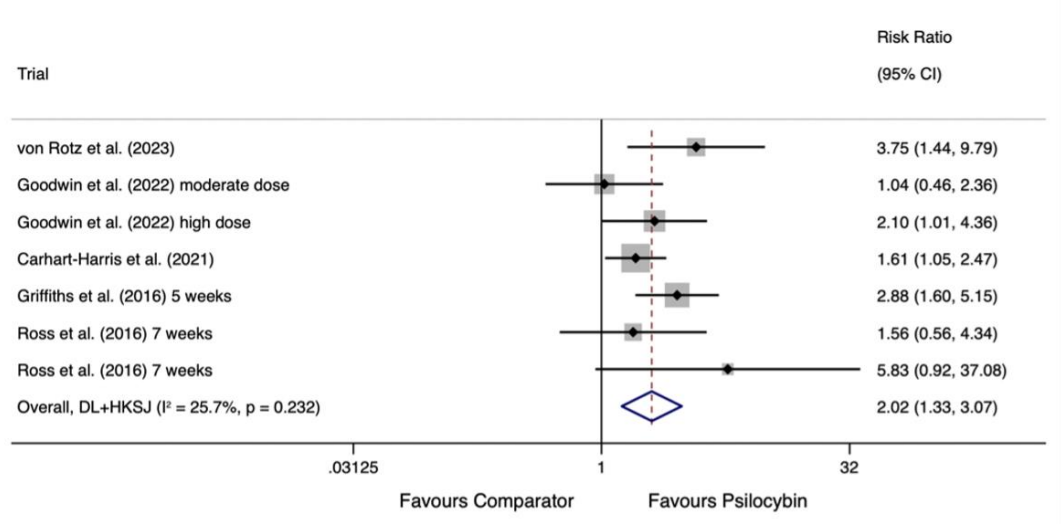


Figure 10. Forest plot for the response rate to treatment by type of depression

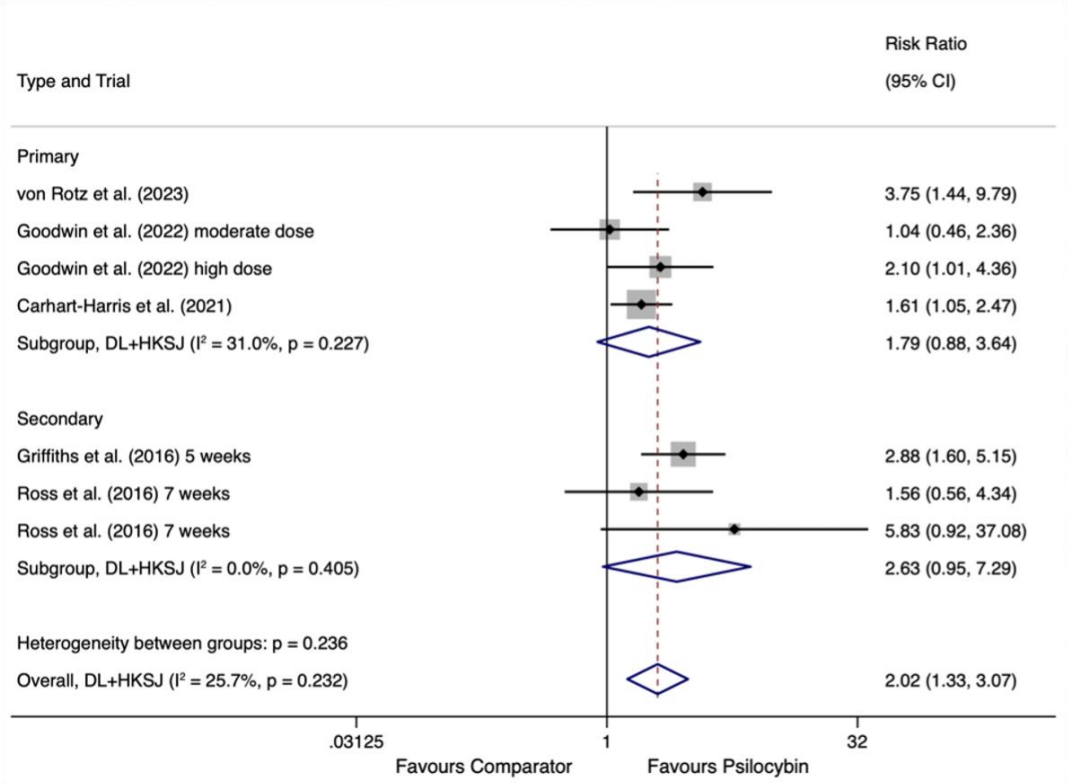


Figure 10. Forest plot for the overall remission rate

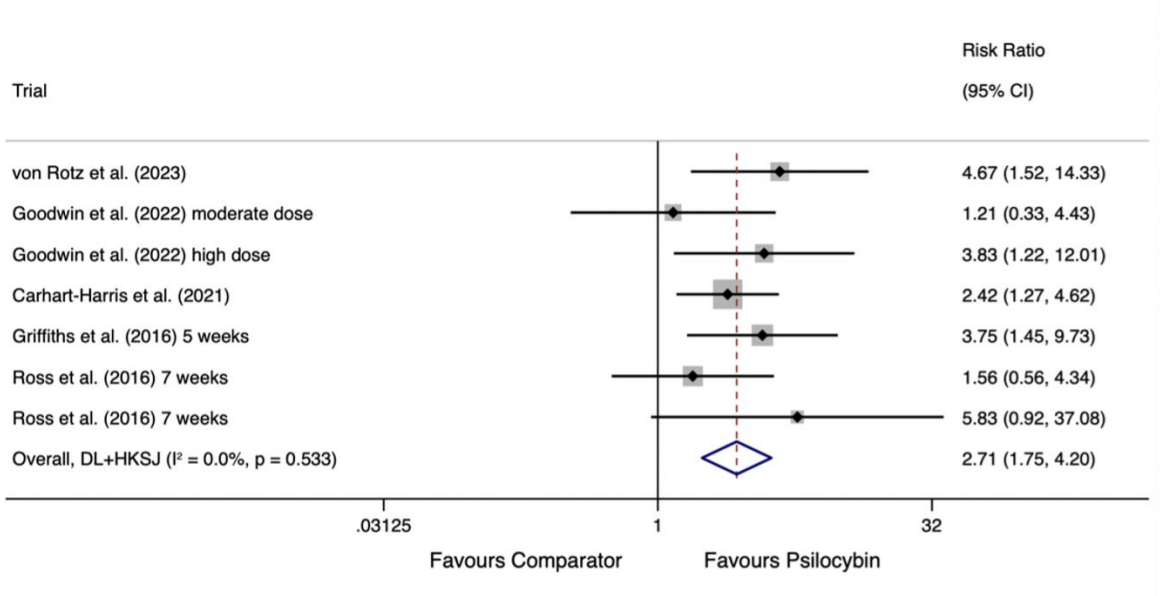


Figure 12. Forest plot for the remission rate by type of depression

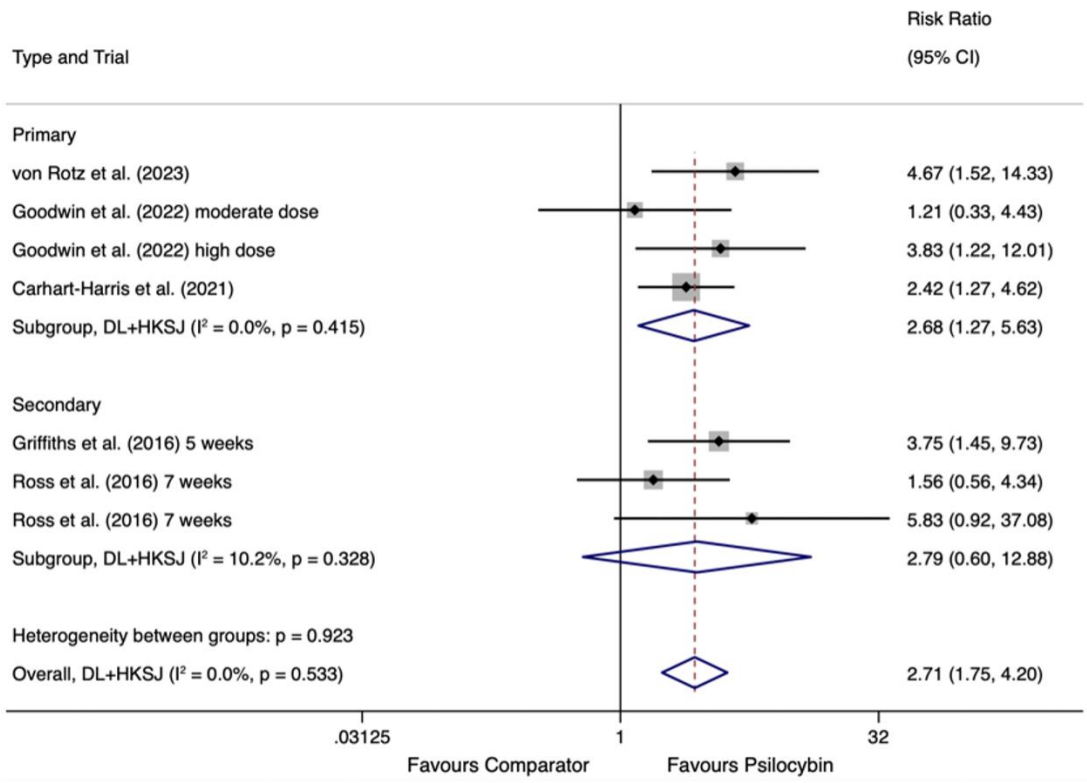
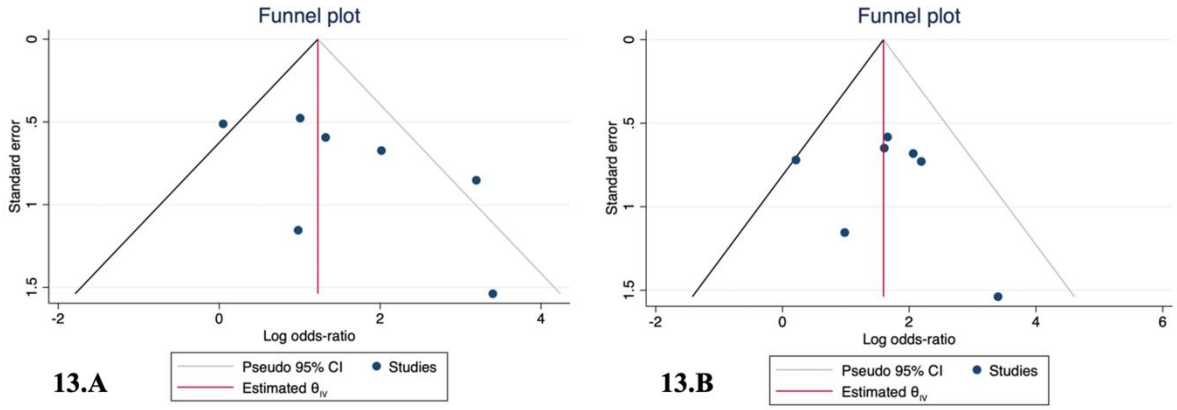


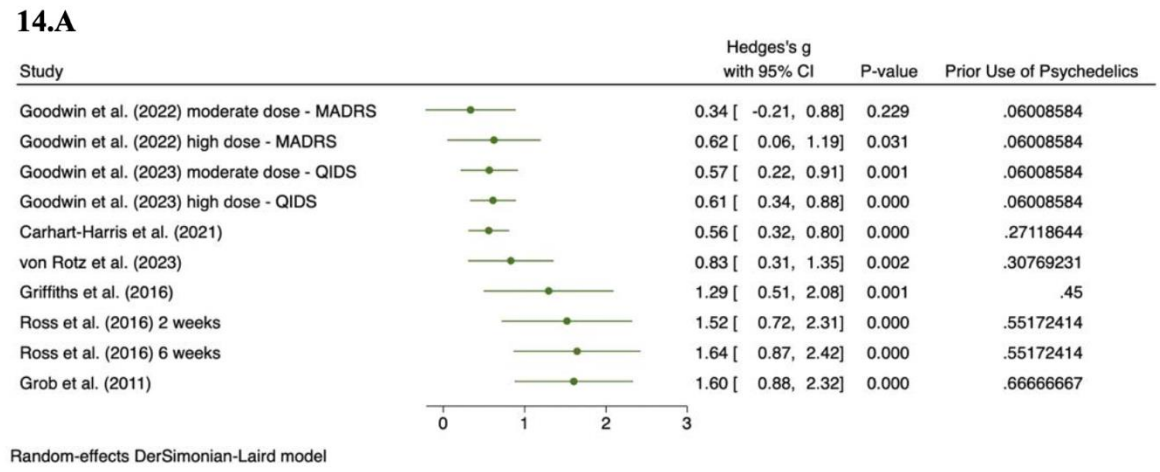
Figure 13. Funnel plot used to assess publication bias among studies measuring (A) treatment response and (B) remission



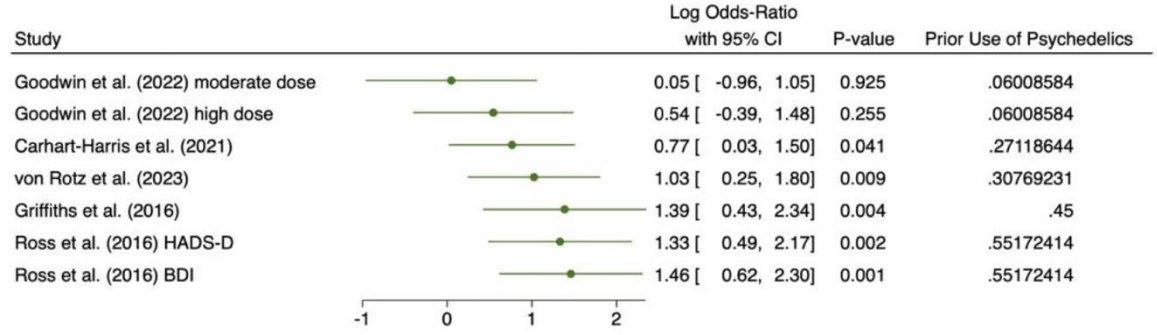
Appendix I: Cumulative meta-analyses

The forest plots show that the change in depression scores (Figure 14A) and likelihood of treatment response (Figure 14B) both increase as the percentage of participants with prior psychedelics use increases across studies, as expected based on the meta-regression analysis. However, participants' age showed a curvilinear relationship with treatment effect (Figure 14C). This is largely due to the breakdown of Goodwin et al.'s (2022; 2023) trials (18, 52) into four separate sets of results (high dose – MADRS, moderate dose – MADRS, high dose – QIDS, moderate dose - QIDS). Thus, the relationship observed between age and treatment effect is likely an artefact of this subdivision, because all other studies show relatively consistent treatment effects with no observable pattern as participants' average age increases (note: Grob et al.'s (2011) study was not included in the meta regression for age, because data on the average age of the participants were not available).

Figure 14. Forest plot of the results of cumulative analyses for the effect of (A) prior psilocybin use on change in depression scores, (B) prior psilocybin use on treatment response, and (C) age on change in depression scores

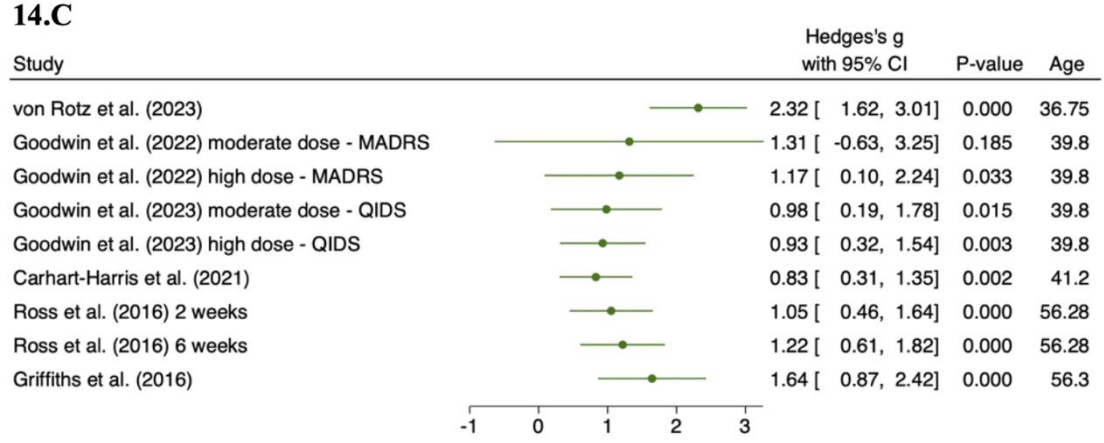


14.B



Random-effects REML model

14.C



Random-effects DerSimonian-Laird model

Appendix J: Sensitivity analysis

The results of the data re-analysis using a random effects Dersimonian and Laird model without the HKSJ modification are seen in Figures 15 and 16. No significant differences from the original model used were found.

To estimate the accuracy and robustness of the estimated treatment effect, studies were excluded 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, both for the continuous outcome (change in depression scores) (Table 3) and dichotomous outcomes (remission and response rates) (Tables 4 and 5), indicating that the results are robust against undue impact by any single study.

Table 3. Results of the sensitivity analysis for changes in depression scores

Removed Study	SMD without this study	Lower limit	Upper limit	Heterogeneity (I^2)	p-value	Prediction Intervals
von Rotz et al. (2023)	1.54	0.32	2.77	89.0%	P<0.001	-1.1 to 4.1
Goodwin et al. (2022) moderate dose	1.82	0.63	3.01	90.1%	P<0.001	-1.1 to 4.6
Goodwin et al. (2022) high dose	1.76	0.52	2.99	90.9%	P<0.001	-1.2 to 4.7
Carhart-Harris et al. (2021)	1.82	0.63	3.01	90.1%	P<0.001	-1.1 to 4.6
Goodwin et al. (2023) moderate dose	1.81	0.61	3.01	90.4%	P<0.001	-1.1 to 4.6
Goodwin et al. (2023) high dose	1.78	0.55	3.00	90.8%	P<0.001	-1.2 to 2.6
Grob et al. (2011)	1.67	0.44	2.90	90.8%	P<0.001	-1.1 to 4.4
Griffiths et al. (2016)	1.23	0.35	2.11	81.9%	P<0.001	-0.7 to 3.1
Ross et al. (2016) 2 weeks	1.46	0.37	2.54	89.7%	P<0.001	-1.1 to 4.0
Ross et al. (2016) 6 weeks	1.52	0.34	2.69	90.2%	P<0.001	-1.1 to 4.1

Table 4. Results of the sensitivity analysis for Response rates

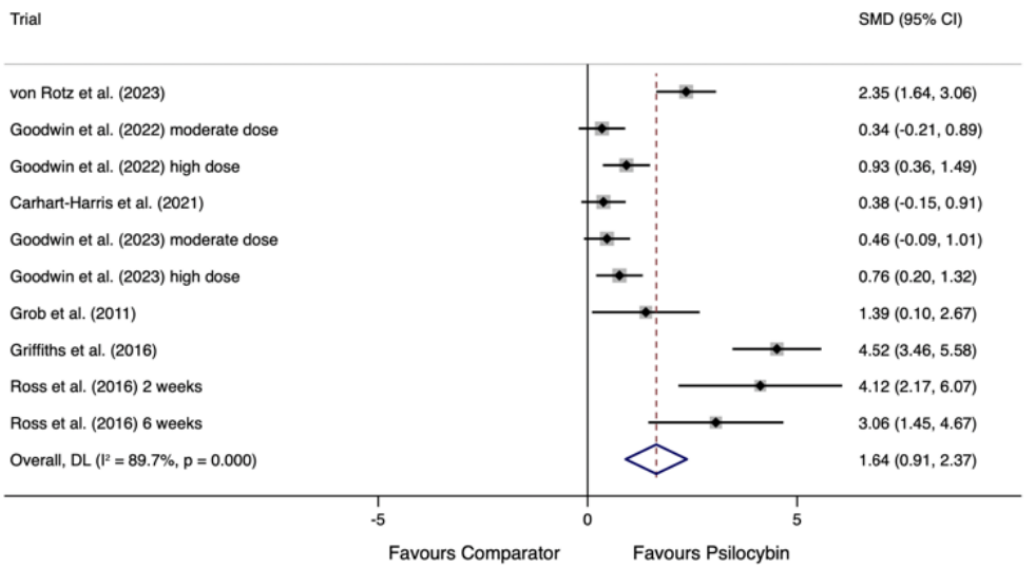
Excluded Study	RR without this study	Lower limit	Upper limit	Heterogeneity (I^2)	p-value	Prediction Intervals
von Rotz et al. (2023)	1.88	1.22	2.90	19.3%	0.288	-1.53 to 4.3
Goodwin et al. (2022) moderate dose	2.16	1.45	3.21	9.0%	0.359	-0.02 to 3.4
Goodwin et al. (2022) high dose	2.03	1.18	3.49	38.7%	0.154	-1.5 to 4.7
Carhart-Harris et al. (2021)	2.21	1.29	3.81	25.5%	0.243	-1.6 to 4.7
Griffiths et al. (2016)	1.82	1.13	2.93	16.9%	0.305	-0.7 to 3.0
Ross et al. (2016) (HADS D)	2.09	1.26	3.48	36.4%	0.164	-1.4 to 4.5
Ross et al. (2016) (BDI)	1.95	1.26	3.01	25.6%	0.242	-1.2 to 3.8

Table 5. Results of the sensitivity analysis for Remission rates

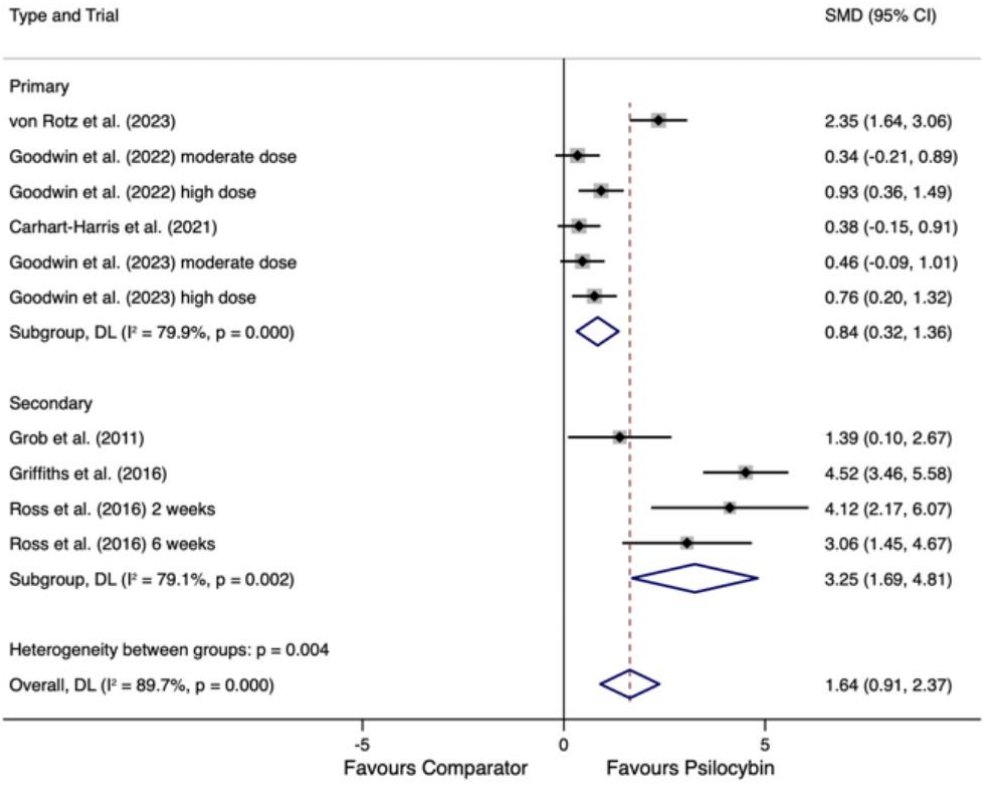
Excluded Study	RR without this study	Lower limit	Upper limit	Heterogeneity (I^2)	p-value	Prediction Intervals
von Rotz et al. (2023)	2.52	1.56	4.08	0.0%	0.540	0.5 to 2.5
Goodwin et al. (2022) moderate dose	2.92	1.89	4.52	0.0%	0.628	1.0 to 2.7
Goodwin et al. (2022) high dose	2.59	1.55	4.33	0.0%	0.456	0.1 to 3.1
Carhart-Harris et al. (2021)	2.88	1.56	5.31	0.0%	0.428	-0.02 to 3.2
Griffiths et al. (2016)	2.55	1.51	4.29	0.0%	0.473	0.3 to 2.6
Ross et al. (2016) (HADS D)	2.96	1.85	4.73	0.0%	0.582	0.6 to 2.7
Ross et al. (2016) (BDI)	2.71	1.75	4.20	0.0%	0.533	0.9 to 2.3

Figure 15. Forest plots for the (A) overall change in depression scores pre- and post-treatment, and change in depression scores by (B) depression type, (C) symptom measurement scale, (D) psilocybin dosage, and (E) time of assessment, using a Dersimonian and Laird model without an HKSJ adjustment

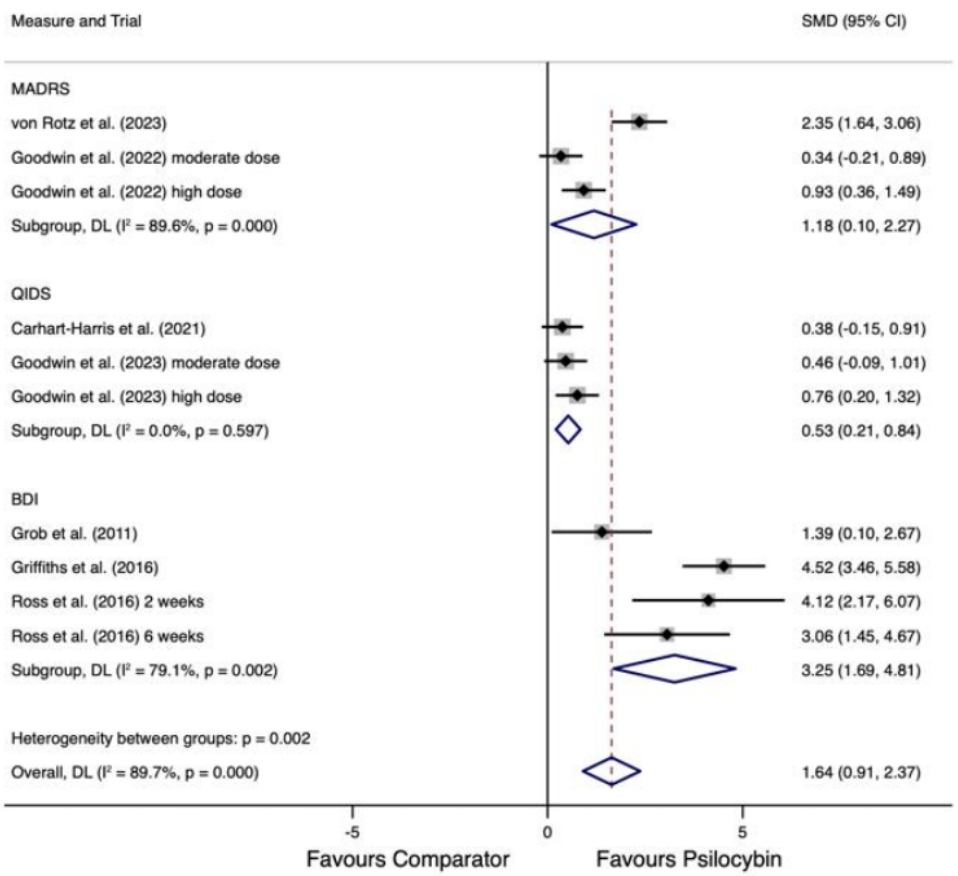
15.A



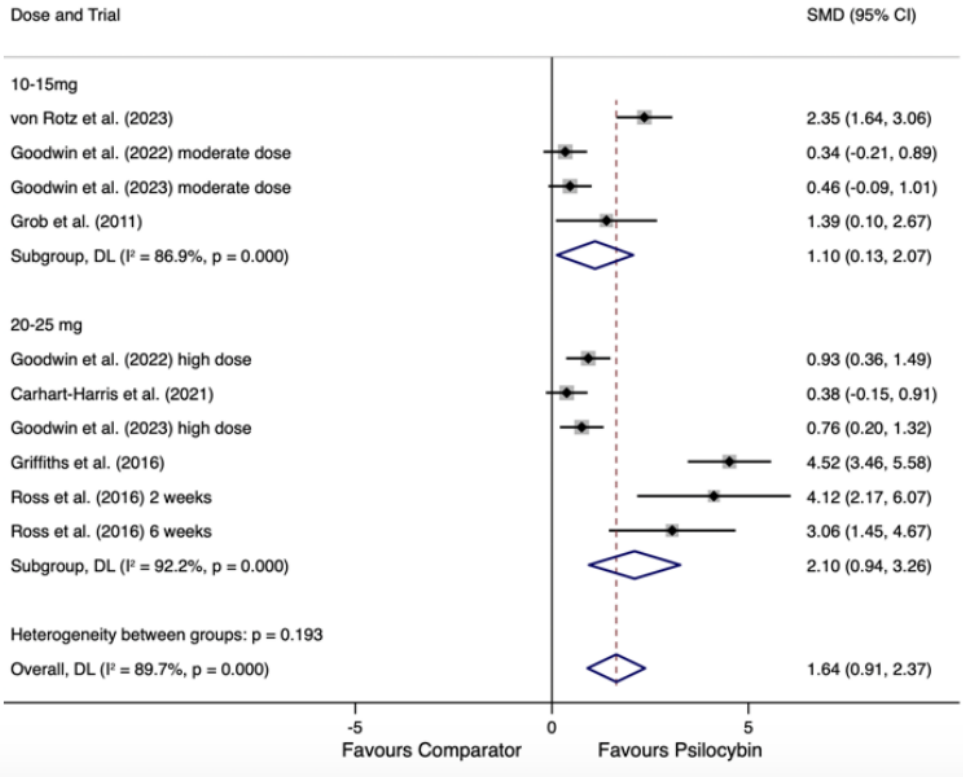
15.B



15.C


Measure and Trial
SMD (95% CI)
MADRS
von Rotz et al. (2023)
2.35 (1.64, 3.06)
Goodwin et al. (2022) moderate dose
0.34 (-0.21, 0.89)
Goodwin et al. (2022) high dose
0.93 (0.36, 1.49)
Subgroup, DL (I² = 89.6%, p = 0.000)
1.18 (0.10, 2.27)
QIDS
Carhart-Harris et al. (2021)
0.38 (-0.15, 0.91)
Goodwin et al. (2023) moderate dose
0.46 (-0.09, 1.01)
Goodwin et al. (2023) high dose
0.76 (0.20, 1.32)
Subgroup, DL (I² = 0.0%, p = 0.597)
0.53 (0.21, 0.84)
BDI
Grob et al. (2011)
1.39 (0.10, 2.67)
Griffiths et al. (2016)
4.52 (3.46, 5.58)
Ross et al. (2016) 2 weeks
4.12 (2.17, 6.07)
Ross et al. (2016) 6 weeks
3.06 (1.45, 4.67)
Subgroup, DL (I² = 79.1%, p = 0.002)
3.25 (1.69, 4.81)
Heterogeneity between groups: p = 0.002
Overall, DL (I² = 89.7%, p = 0.000)
1.64 (0.91, 2.37)
-5
0
5
Favours Comparator
Favours Psilocybin

15.D


Dose and Trial
SMD (95% CI)
10-15mg
von Rotz et al. (2023)
2.35 (1.64, 3.06)
Goodwin et al. (2022) moderate dose
0.34 (-0.21, 0.89)
Goodwin et al. (2023) moderate dose
0.46 (-0.09, 1.01)
Grob et al. (2011)
1.39 (0.10, 2.67)
Subgroup, DL (I² = 86.9%, p = 0.000)
1.10 (0.13, 2.07)
20-25 mg
Goodwin et al. (2022) high dose
0.93 (0.36, 1.49)
Carhart-Harris et al. (2021)
0.38 (-0.15, 0.91)
Goodwin et al. (2023) high dose
0.76 (0.20, 1.32)
Griffiths et al. (2016)
4.52 (3.46, 5.58)
Ross et al. (2016) 2 weeks
4.12 (2.17, 6.07)
Ross et al. (2016) 6 weeks
3.06 (1.45, 4.67)
Subgroup, DL (I² = 92.2%, p = 0.000)
2.10 (0.94, 3.26)
Heterogeneity between groups: p = 0.193
Overall, DL (I² = 89.7%, p = 0.000)
1.64 (0.91, 2.37)
-5
0
5
Favours Comparator
Favours Psilocybin

15.E

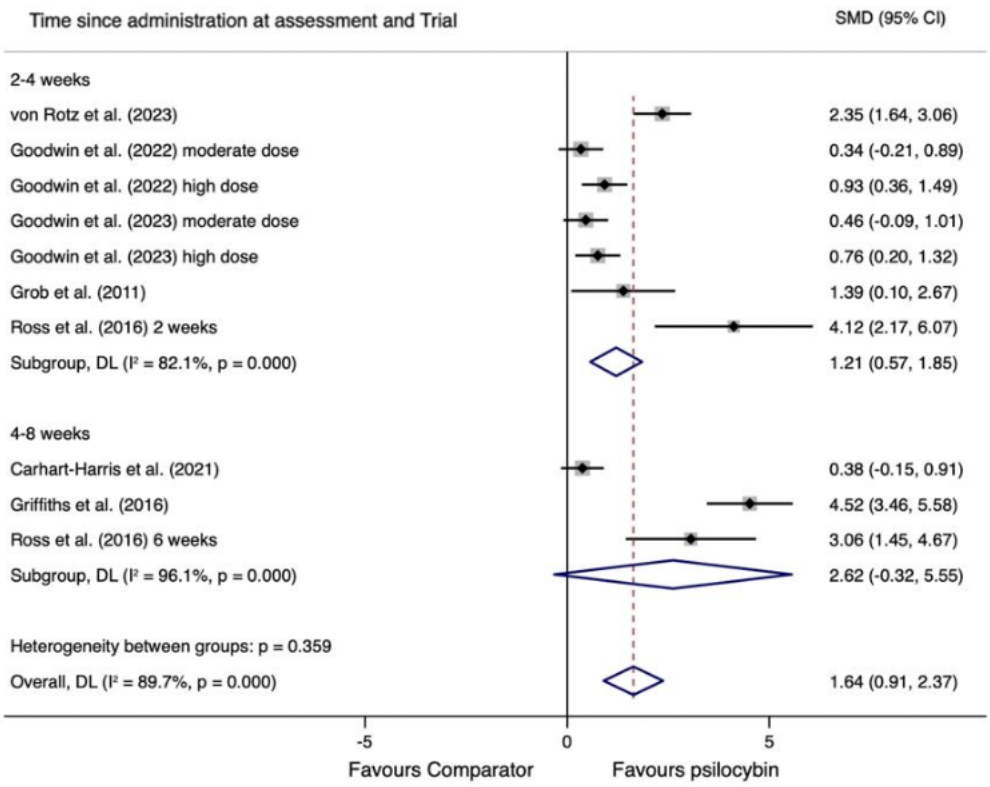
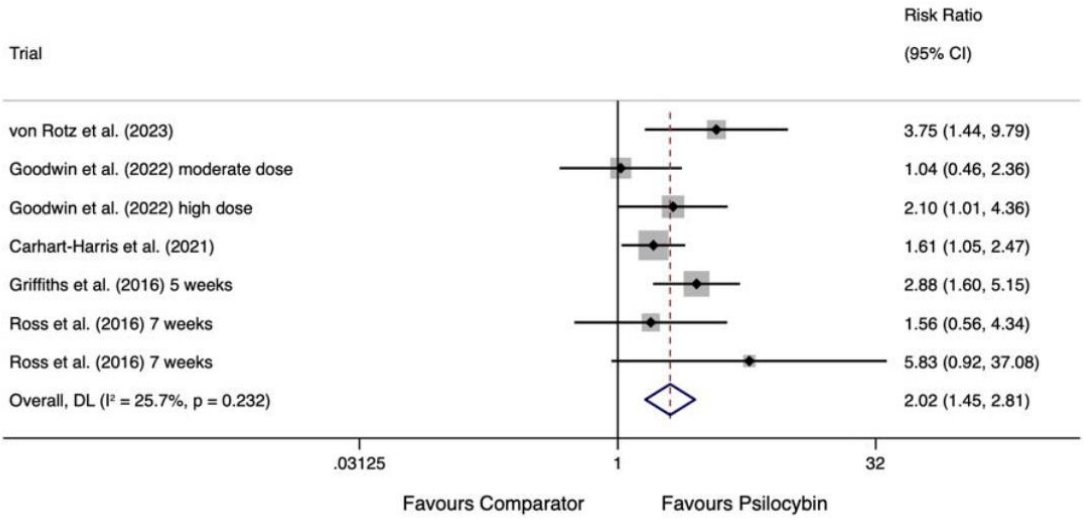
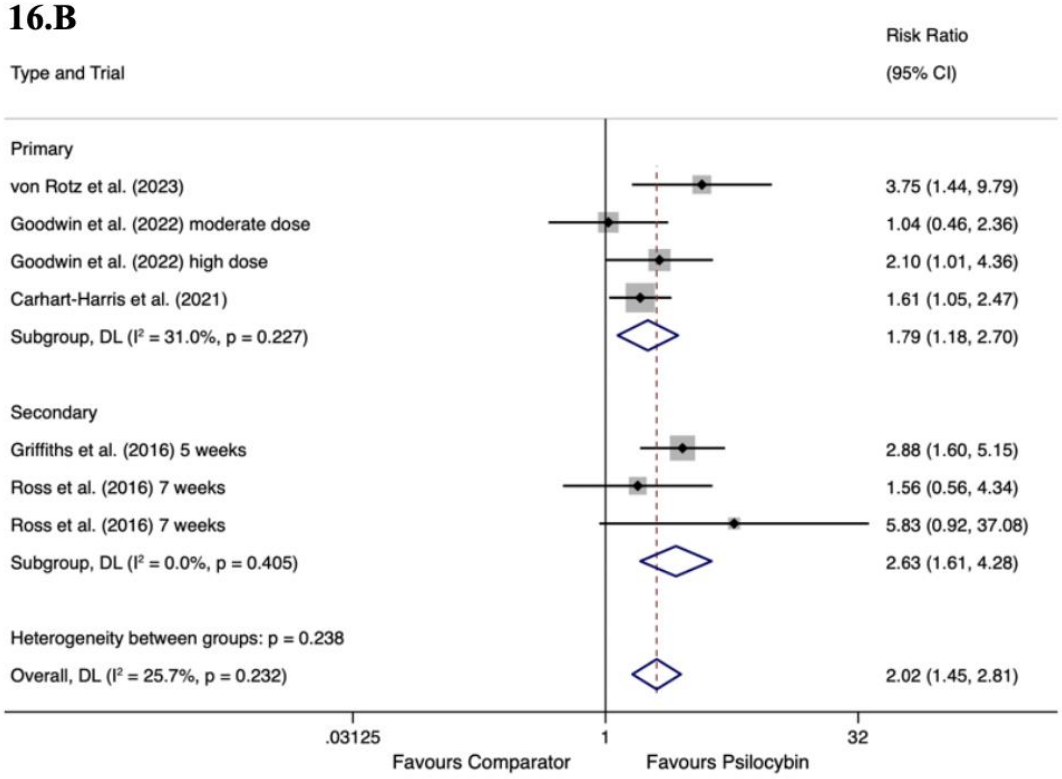


Figure 16. Forest plots for the (A) overall change in response rates, (B) change in response rates by depression type, (C) overall change in remission rates, and (D) change in remission rates by depression type, using a Dersimonian and Laird model without an HKSJ adjustment

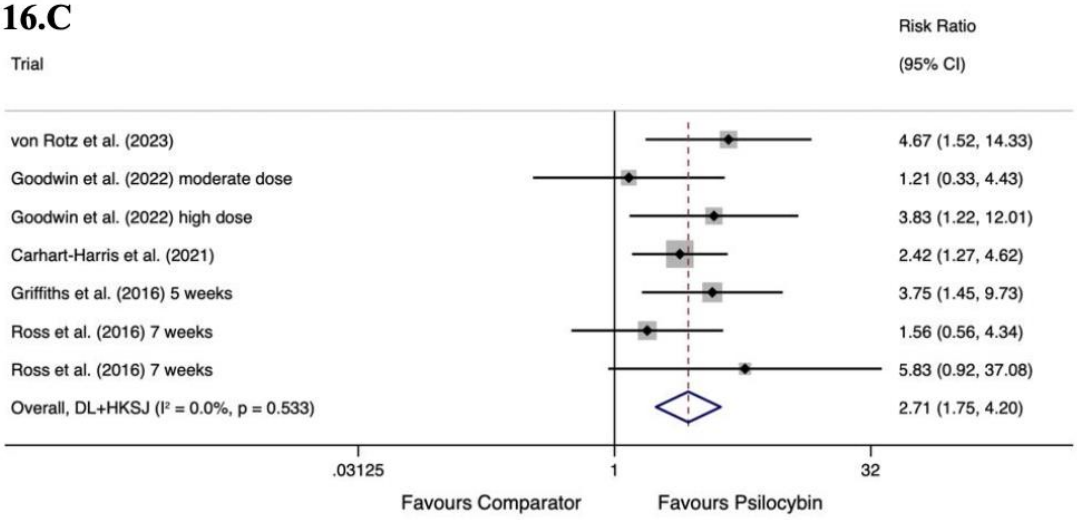
16.A



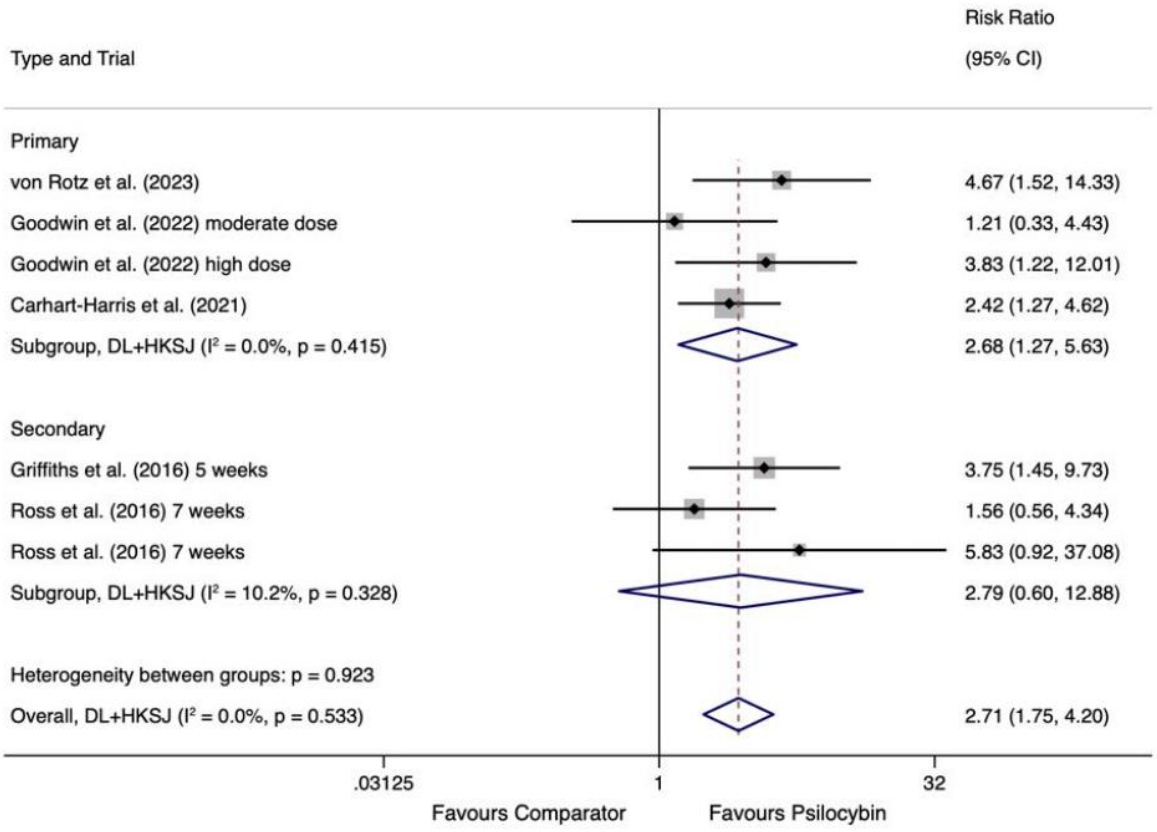
16.B


Type and Trial
Risk Ratio
(95% CI)
Primary
von Rotz et al. (2023)
3.75 (1.44, 9.79)
Goodwin et al. (2022) moderate dose
1.04 (0.46, 2.36)
Goodwin et al. (2022) high dose
2.10 (1.01, 4.36)
Carhart-Harris et al. (2021)
1.61 (1.05, 2.47)
Subgroup, DL (I² = 31.0%, p = 0.227)
1.79 (1.18, 2.70)
Secondary
Griffiths et al. (2016) 5 weeks
2.88 (1.60, 5.15)
Ross et al. (2016) 7 weeks
1.56 (0.56, 4.34)
Ross et al. (2016) 7 weeks
5.83 (0.92, 37.08)
Subgroup, DL (I² = 0.0%, p = 0.405)
2.63 (1.61, 4.28)
Heterogeneity between groups: p = 0.238
Overall, DL (I² = 25.7%, p = 0.232)
2.02 (1.45, 2.81)
.03125
1
32
Favours Comparator
Favours Psilocybin

16.C


Trial
Risk Ratio
(95% CI)
von Rotz et al. (2023)
4.67 (1.52, 14.33)
Goodwin et al. (2022) moderate dose
1.21 (0.33, 4.43)
Goodwin et al. (2022) high dose
3.83 (1.22, 12.01)
Carhart-Harris et al. (2021)
2.42 (1.27, 4.62)
Griffiths et al. (2016) 5 weeks
3.75 (1.45, 9.73)
Ross et al. (2016) 7 weeks
1.56 (0.56, 4.34)
Ross et al. (2016) 7 weeks
5.83 (0.92, 37.08)
Overall, DL+HKSJ (I² = 0.0%, p = 0.533)
2.71 (1.75, 4.20)
.03125
1
32
Favours Comparator
Favours Psilocybin

16.D


Risk Ratio
Type and Trial
(95% CI)
Primary
von Rotz et al. (2023)
4.67 (1.52, 14.33)
Goodwin et al. (2022) moderate dose
1.21 (0.33, 4.43)
Goodwin et al. (2022) high dose
3.83 (1.22, 12.01)
Carhart-Harris et al. (2021)
2.42 (1.27, 4.62)
Subgroup, DL+HKSJ (I² = 0.0%, p = 0.415)
2.68 (1.27, 5.63)
Secondary
Griffiths et al. (2016) 5 weeks
3.75 (1.45, 9.73)
Ross et al. (2016) 7 weeks
1.56 (0.56, 4.34)
Ross et al. (2016) 7 weeks
5.83 (0.92, 37.08)
Subgroup, DL+HKSJ (I² = 10.2%, p = 0.328)
2.79 (0.60, 12.88)
Heterogeneity between groups: p = 0.923
Overall, DL+HKSJ (I² = 0.0%, p = 0.533)
2.71 (1.75, 4.20)
.03125
1
32
Favours Comparator
Favours Psilocybin