

Assessing the risk of symptom worsening in psilocybin-assisted therapy for depression: a systematic review and individual participant data meta-analysis

Otto Simonsson^{1,2}, Per Carlbring³, Robin Carhart-Harris^{4,5}, Alan K. Davis^{6,7,8}, David J. Nutt⁵, Roland R. Griffiths^{8,9,10}, David Erritzoe⁵, Simon B. Goldberg¹¹

¹Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden

²Department of Sociology, University of Oxford, Oxford, UK

³Department of Psychology, Stockholm University, Stockholm, Sweden

⁴Neuroscape Psychedelics Division, Department of Neurology, University of California San Francisco, San Francisco, CA, United States

⁵Centre for Psychedelic Research, Imperial College London, London, UK

⁶Center for Psychedelic Drug Research and Education, College of Social Work, The Ohio State University, Columbus, OH, USA

⁷Department of Psychiatry, The Ohio State University, Columbus, OH, USA

⁸Center for Psychedelic and Consciousness Research, Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA

⁹Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

¹⁰Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

¹¹Department of Counseling Psychology, University of Wisconsin, Madison, WI, USA

Corresponding Author Contact Details

Otto Simonsson, otto.simonsson@ki.se
Norra Stationsgatan 69, 113 64 Stockholm, Sweden

ORCID numbers

Otto Simonsson: 0000-0003-4197-7566
Per Carlbring: 0000-0002-2172-8813
Robin Carhart-Harris: 0000-0002-6062-7150
Alan K. Davis: 0000-0003-4770-8893
David J. Nutt: 0000-0002-1286-1401
David Erritzoe: 0000-0002-7022-6211
Simon B. Goldberg: 0000-0002-6888-0126

Keywords: Psilocybin; depression; meta-analysis

1. Introduction

The leading factor contributing to disability worldwide is depression, a mood disorder that is estimated to affect more than 350 million people (Cuijpers et al., 2020a). Standard treatments for depression are effective for some patients, but many do not respond to treatment at all, and some experience worsening of depressive symptoms (Kolovos et al., 2017). Such treatments can also take weeks or even months to produce clinically relevant reductions in depressive symptoms, highlighting the need for novel treatments for depressive disorders (Cuijpers et al., 2020b).

One intervention that shows promise is psilocybin-assisted therapy (Nutt & Carhart-Harris, 2021). The administration of psilocybin in conjunction with therapy has been shown to reduce depressive symptoms in several clinical trials (Leger & Unterwald, 2022), but no study to date has evaluated clinically relevant worsening of depressive symptoms in psilocybin clinical trials for depression. There is also limited information on whether baseline demographic characteristics are associated with symptom worsening or treatment response to psilocybin-assisted therapy (Aday et al., 2021).

In this study, we identified all published psilocybin clinical trials on depression. We requested the primary depression outcome data from study authors and conducted an individual participant data meta-analysis 1) assessing prevalence of clinically relevant worsening of depressive symptoms and 2) examining baseline demographic characteristics associated with symptom worsening or treatment response. We hypothesized that rates of clinically relevant worsening of depressive symptoms would be lower in psilocybin conditions than rates in control conditions for studies that included control groups, but we had no *a priori* hypotheses about baseline demographic characteristics associated with symptom worsening or treatment response.

2. Methods

This independent participant data meta-analysis is reported following the PRISMA guidelines (Stewart et al., 2015). The study protocols were registered at the Open Science Framework: <https://osf.io/ctfzs> and <https://osf.io/jwbkf>. Deviations from our preregistration are reported in Supplemental Materials. The study was determined to be exempt from review by the Internal Review Board (IRB) at UW-Madison.

2.1 Search Strategy and Study Selection

We searched PubMed, PsycINFO, Embase and the Cochrane Library with the following search term: psilo*. The search was conducted on 28th March, 2022. The databases were searched since their inception. No restrictions were placed on language or publication status. Studies that had this term appear in the abstract, title, and/or keywords were reviewed. Bibliographies of recent meta-analyses examining psilocybin and psychedelic trials were also searched for potentially relevant studies (Li et al., 2022; Kisely et al., 2022; Leger & Unterwald, 2022; Yu et al., 2022; Zeifman et al., 2022). Eligible studies had to have used psilocybin as the primary intervention and have reported outcome data on standardized measures of depression. Controlled and uncontrolled studies on both clinical and non-clinical populations were eligible (see Supplemental Materials for information about data extraction).

2.2 Statistical analyses

To characterize symptom change, we calculated standardized mean difference (SMD) scores for the depression measures (GRID-Hamilton Depression Rating Scale; Williams et al., 2008; Quick Inventory of Depression Symptoms Self-Report – 16; Rush et al., 2003), in keeping with meta-analytic methods (Borenstein et al., 2009). Specifically, we calculated pre-post change scores (post minus pre) and divided this value by the baseline standard deviation of each measure. To define symptom worsening, we used a value of $SMD \geq 0.24$ (Cuijpers et al.,

2014). We then conducted a series of one-step meta-analyses with a random effects component (i.e., random intercept multilevel models; Burke et al., 2017) examining predictors of symptom worsening and treatment response. In keeping with Burke and colleagues (2017), we modeled the nesting of effects within study ID. We examined five demographic variables which were available across all three studies as predictors.

Models examining response to psilocybin included the psilocybin arm from all three trials. Models examining treatment response as a continuous variable (SMD) used multilevel linear regression while models examining treatment response as a dichotomous variable (i.e., symptom worsening) used multilevel logistic regression. Analyses were conducted in R (R Core Team, 2022; see Supplemental Materials for R code).

3. Results

Three studies were included in the independent participant data meta-analysis (Carhart-Harris et al., 2016, 2021; Davis et al., 2021; see Fig. 1),¹ which were all of the eligible studies with two dosing sessions focused on populations with depressive disorders (see Supplemental Materials and Supplemental Tables 1-3 for details of studies included). Collectively, these studies included 102 participants who completed the measures at both baseline and at six-week follow-up, of whom 62 received psilocybin-assisted therapy, 29 received escitalopram, and 11 received waitlist. Five baseline demographic characteristics across the three studies were included: age, gender (coded as male versus female), race/ethnicity (coded as White versus non-White), education (coded as undergraduate degree or higher versus other), and employment status (coded as unemployed versus other).

Participants in the psilocybin and escitalopram conditions showed large reductions in depressive symptoms at post-test in both conditions (SMDs = -2.38 and -1.56, SD = 1.69 and

¹ The full sample from Carhart-Harris and colleagues (2016) is reported in Carhart-Harris and colleagues(2018).

1.36, respectively) while participants in the waitlist control showed a worsening of symptoms on average (SMD = 0.26, SD = 1.06). A minority of participants in the psilocybin and escitalopram conditions showed clinically significant symptom worsening (9.7% and 10.3%, respectively), while the majority of participants in the waitlist control condition showed clinically significant symptom worsening (63.6%; see Supplemental Table 4). When restricted to the two studies that included a control condition, assignment to the psilocybin arm was associated with a lower likelihood of symptom worsening relative to waitlist (OR = 13.30, 95% CI [3.02, 70.74], $p = .001$) and no difference in the likelihood of symptom worsening relative to escitalopram (OR = 0.88, 95% CI [0.17, 3.89], $p = .865$).²

None of the five demographic variables examined were associated with response to the psilocybin arm (Supplemental Table 5). One empty cell was detected when examining demographic variables in association with symptom worsening. Specifically, no non-White participants reported worsening symptoms following psilocybin. To examine this demographic predictor, we implemented Firth's (1993) bias-reduced penalized likelihood logistic regression implemented in the 'logistf' package in R (Heinze, Ploner & Jiricka, 2022). None of the five demographics variables examined were associated with likelihood of symptom worsening in response to psilocybin (Supplemental Table 6).³

4. Discussion

This study was an individual participant data meta-analysis assessing the prevalence of clinically relevant worsening of depressive symptoms and examining baseline demographic characteristics associated with symptom worsening or treatment response. Results showed

² Significance tests did not change when including the five demographic variables as covariates (OR = 35.83, 95% CI [5.33, 407.26], $p < .001$ for psilocybin vs. waitlist; OR = 1.20, 95% CI [0.19, 7.16], $p = .833$ for psilocybin vs. escitalopram).

³ Whether the study focused on participants with treatment-resistant depression was not associated with frequency of symptom worsening (5.3% for treatment-resistant depression vs. 18.1% for other studies, OR = 0.42, 95% CI [0.02, 3.30], $p = .446$).

clinically significant symptom worsening in a minority (~10%) of participants in the psilocybin and escitalopram conditions. This is in line with rates for psychotherapy, where ~7% of the patients show symptom worsening (Mechler & Holmqvist, 2016). By contrast, a majority (63.6%) of the waitlist condition showed symptom worsening. That is a surprisingly high proportion when compared with a meta-analysis of waitlist controls in psychotherapy that found only 17.4% of patients showed symptom worsening (Rozental et al., 2017). However, had the psychotherapy meta-analysis used the same conservative cut-off of 0.24 instead of 0.84 SMD units, the proportion may have been comparable. This relatively high rate of worsening in the waitlist condition may reflect a kind of “nocebo” effect, where participants not receiving a desired treatment are actively disappointed, resulting in symptom worsening. Worsening associated with waitlist conditions specifically has been observed in psychotherapy trials previously (Furukawa et al., 2014).

In the two clinical trials with control conditions, assignment to the psilocybin arm was associated with a lower likelihood of symptom worsening relative to waitlist and no difference in the likelihood of symptom worsening relative to escitalopram. Thus, it appears that receipt of psilocybin confers risk of symptom worsening similar to an FDA-approved antidepressant medication and is substantially protective against risk of symptom worsening relative to treatment with delayed start (i.e., waitlist; Cuijpers & Cristea, 2016). None of the five baseline demographic characteristics examined were associated with response to psilocybin or likelihood of symptom worsening in response to psilocybin.

There are several limitations to consider when interpreting the results of this study. First, the combined sample size of the included studies was relatively small, which limited statistical power to detect potentially smaller magnitude associations. The sample size of the waitlist control condition (n=11) was especially small and may therefore have impacted the reliability of comparisons. Second, there are many ways to operationalize worsening of clinical status

(e.g., increase in suicidality), but this study focused solely on worsening of depressive symptoms. Third, the included studies were heterogeneous in terms of research design. Fourth, participant-level predictors were limited to five baseline demographic characteristics. It would be useful in future studies to examine additional potential predictors of treatment response (e.g., psychological, genetic). Fifth, the diversity (e.g., race and ethnicity) in the samples was limited and should be addressed in future studies to increase the generalizability of findings (Michaels et al., 2018). Sixth, only six-week follow-up was examined in this study. It was therefore not possible for this analysis to provide guidance on the time course of symptom worsening or any sustained effects beyond these assessments.

Although the findings in this study should be considered preliminary, these results suggest that clinically relevant symptom worsening in depressed patients is not more common with psilocybin-assisted therapy than with standard pharmacological treatment (i.e., escitalopram). If such findings are replicated in future studies, it would further strengthen the overall safety profile of psilocybin, which appears favorable based on the evidence to date (Roscoe & Lozy, 2022).

Author contributions

OS, PC, and SBG conceptualized and preregistered the study. OS and PC conducted the screening. SBG supervised the study and conducted the analyses. RCH, AKD, DJN, RRG, DE provided data and made critical revisions.

Funding

OS was supported by Ekhaga Foundation and Olle Engkvist Foundation. SBG was supported by the National Center for Complementary & Integrative Health of the National Institutes of Health under Award Number K23AT010879. AKD is supported by the Center for Psychedelic Drug Research and Education. RRG and AKD are supported by the Johns

Hopkins Center for Psychedelic and Consciousness Research with funding from Tim Ferriss, Matt Mullenweg, Craig Nerenberg, Blake Mycoskie, the Steven and Alexandra Cohen Foundation.

Declaration of competing interest

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: OS was a co-founder of Eudelics AB. RCH is a scientific advisor to Synthesis Institute, Osmind, Journey Colab, Maya Health, Mydecine, Beckley Psytech and Mindstate. AKD is a board member of Source Research Foundation and Lead Training at Fluence. DJN is a scientific advisor to COMPASS Pathways who have an interest in psilocybin therapy for depression. He is also chair of PAREA (Psychedelic access and research European alliance) and a member of the UK Drug Science charity's Medical Psychedelic Working Group. DE is a scientific advisor to Clerkenwell Health, Aya Biosciences, Field Trip Health, Mindstate, Pangea Botanica, and Smallpharma LTD. RRG is on the Board of Directors of the Heffter Research Institute. All other authors declare that there is no conflict of interest.

Acknowledgments

The authors would like to thank Joseph M Peill for his help in transferring the independent participant data from the trials at Imperial College London. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Aday, J. S., Davis, A. K., Mitzkovitz, C. M., Bloesch, E. K., & Davoli, C. C. (2021). Predicting reactions to psychedelic drugs: A systematic review of states and traits related to acute drug effects. *ACS Pharmacology & Translational Science*, 4(2), 424-435.

Borenstein, M., Cooper, H., Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. *The handbook of research synthesis and meta-analysis*, 2, 221-235.

Burke, D. L., Ensor, J., & Riley, R. D. (2017). Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Statistics in medicine*, 36(5), 855-875.

Carhart-Harris, R. L., Bolstridge, M., Rucker, J., Day, C. M., Erritzoe, D., Kaelen, M., ... & Nutt, D. J. (2016). Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *The Lancet Psychiatry*, 3(7), 619-627.

Carhart-Harris, R. L., Bolstridge, M., Day, C. M., Rucker, J., Watts, R., Erritzoe, D. E., ... & Nutt, D. J. (2018). Psilocybin with psychological support for treatment-resistant depression: six-month follow-up. *Psychopharmacology*, 235, 399-408.

Carhart-Harris, R., Giribaldi, B., Watts, R., Baker-Jones, M., Murphy-Beiner, A., Murphy, R., ... & Nutt, D. J. (2021). Trial of psilocybin versus escitalopram for depression. *New England Journal of Medicine*, 384(15), 1402-1411.

Cuijpers, P., & Cristea, I. A. (2016). How to prove that your therapy is effective, even when it is not: a guideline. *Epidemiology and Psychiatric Sciences*, 25(5), 428-435.

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, A., & Furukawa, T. A. (2020b). A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*, 19(1), 92-107.

Cuijpers, P., Stringaris, A., & Wolpert, M. (2020a). Treatment outcomes for depression: challenges and opportunities. *The Lancet Psychiatry*, 7(11), 925-927.

Cuijpers, P., Turner, E. H., Koole, S. L., Van Dijke, A., & Smit, F. (2014). What is the threshold for a clinically relevant effect? The case of major depressive disorders. *Depression and anxiety*, 31(5), 374-378.

Davis, A. K., Barrett, F. S., May, D. G., Cosimano, M. P., Sepeda, N. D., Johnson, M. W., ... & Griffiths, R. R. (2021). Effects of psilocybin-assisted therapy on major depressive disorder: a randomized clinical trial. *JAMA psychiatry*, 78(5), 481-489.

Firth D (1993). Bias reduction of maximum likelihood estimates. *Biometrika* 80, 27-38.

Furukawa, T. A., Noma, H., Caldwell, D. M., Honyashiki, M., Shinohara, K., Imai, H., ... & Churchill, R. (2014). Waiting list may be a nocebo condition in psychotherapy trials: A contribution from network meta-analysis. *Acta Psychiatrica Scandinavica*, 130(3), 181-192.

Heinze, G., Ploner, M., Jiricka, L. (2022). logistf: Firth's Bias-Reduced Logistic Regression. R package version 1.24.1, <<https://CRAN.R-project.org/package=logistf>>.

Kisely, S., Connor, M., Somogyi, A. A., & Siskind, D. (2022). A systematic literature review and meta-analysis of the effect of psilocybin and methylenedioxymethamphetamine on mental, behavioural or developmental disorders. *Australian & New Zealand Journal of Psychiatry*, 00048674221083868.

Kolovos, S., van Tulder, M. W., Cuijpers, P., Prigent, A., Chevreur, K., Riper, H., & Bosmans, J. E. (2017). The effect of treatment as usual on major depressive disorder: a meta-analysis. *Journal of Affective Disorders*, 210, 72-81.

Leger, R. F., & Unterwald, E. M. (2022). Assessing the effects of methodological differences on outcomes in the use of psychedelics in the treatment of anxiety and depressive disorders: A systematic review and meta-analysis. *Journal of Psychopharmacology*, 36(1), 20-30.

Li, N. X., Hu, Y. R., Chen, W. N., & Zhang, B. (2022). Dose effect of psilocybin on primary and secondary depression: a preliminary systematic review and meta-analysis. *Journal of Affective Disorders*, 296, 26-34.

Mechler, J., & Holmqvist, R. (2016). Deteriorated and unchanged patients in psychological treatment in Swedish primary care and psychiatry. *Nordic Journal of Psychiatry*, 70(1), 16-23.

Michaels, T. I., Purdon, J., Collins, A., & Williams, M. T. (2018). Inclusion of people of color in psychedelic-assisted psychotherapy: A review of the literature. *BMC psychiatry*, 18(1), 1-14.

Nutt, D., & Carhart-Harris, R. (2021). The current status of psychedelics in psychiatry. *JAMA psychiatry*, 78(2), 121-122.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Roscoe, J., & Lozy, O. (2022). Can psilocybin be safely administered under medical supervision? A systematic review of adverse event reporting in clinical trials. *Drug Science, Policy and Law*, 8, 20503245221085222.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., ... & Keller, M. B. (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological psychiatry*, 54(5), 573-583.

Rozental, A., Magnusson, K., Boettcher, J., Andersson, G., & Carlbring, P. (2017). For better or worse: An individual patient data meta-analysis of deterioration among participants receiving Internet-based cognitive behavior therapy. *Journal of consulting and clinical psychology*, 85(2), 160.

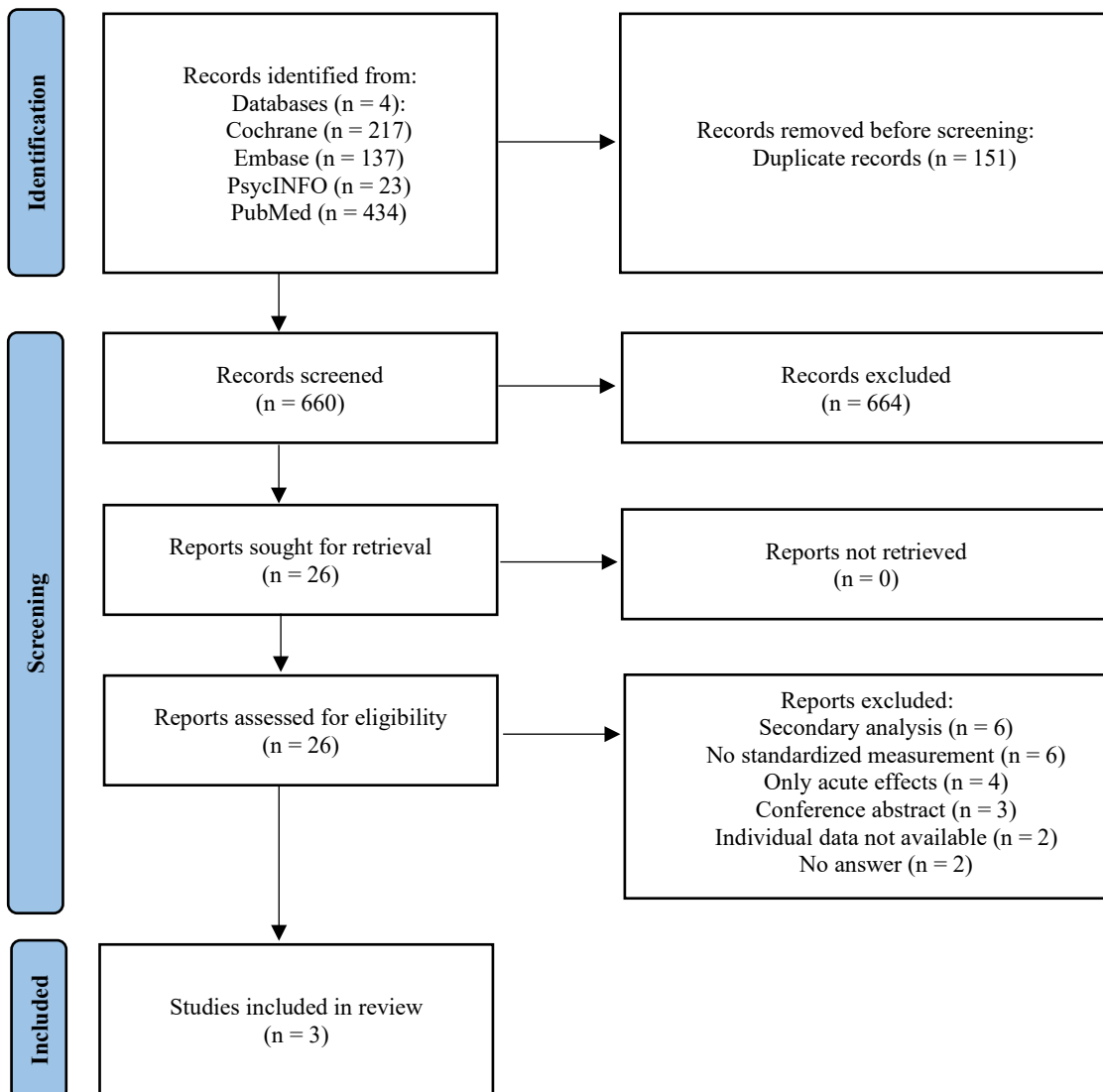
Stewart, L. A., Clarke, M., Rovers, M., Riley, R. D., Simmonds, M., Stewart, G., & Tierney, J. F. (2015). Preferred reporting items for a systematic review and meta-analysis of individual participant data: the PRISMA-IPD statement. *Jama*, 313(16), 1657-1665.

Yu, C. L., Liang, C. S., Yang, F. C., Tu, Y. K., Hsu, C. W., Carvalho, A. F., ... & Su, K. P. (2022). Trajectory of antidepressant effects after single-or two-dose administration of psilocybin: a systematic review and multivariate meta-analysis. *Journal of clinical medicine*, 11(4), 938.

Williams, J. B., Kobak, K. A., Bech, P., Engelhardt, N., Evans, K., Lipsitz, J., ... & Kalali, A. (2008). The GRID-HAMD: standardization of the Hamilton depression rating scale. *International clinical psychopharmacology*, 23(3), 120-129.

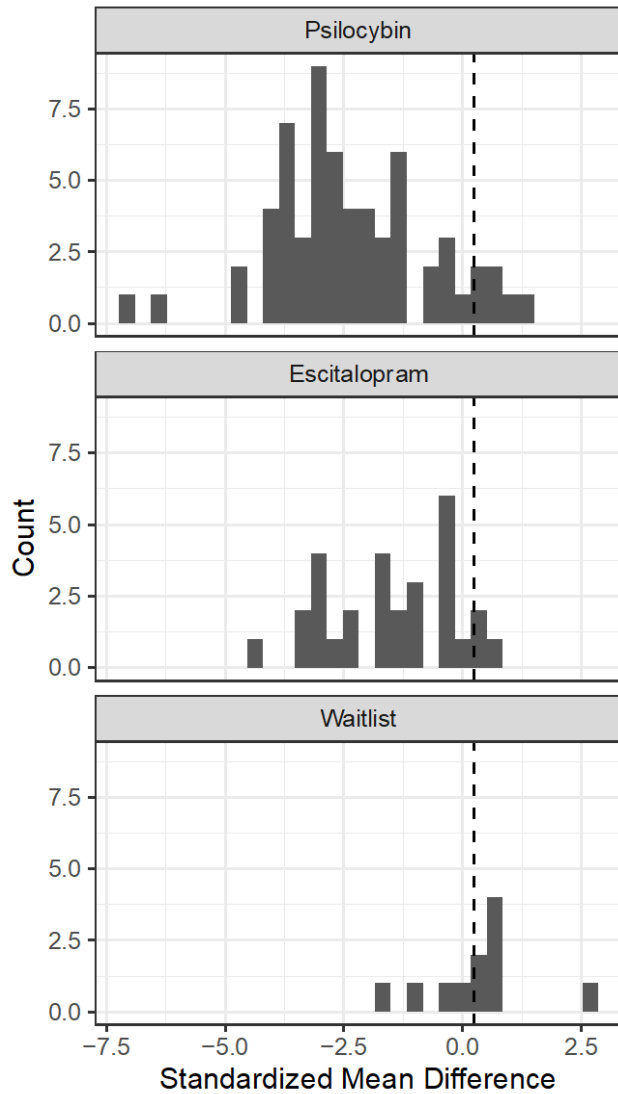
Zeifman, R. J., Yu, D., Singhal, N., Wang, G., Nayak, S. M., & Weissman, C. R. (2022). Decreases in suicidality following psychedelic therapy: a meta-analysis of individual patient data across clinical trials. *The Journal of Clinical Psychiatry*, 83(2), 39235

Figure 1



Supplemental Figure 1

Histogram of pre-post standardized mean differences (SMD) separated by treatment group. Dashed line at SMD = 0.24 reflecting clinically significant worsening of depression symptoms.



Supplemental Table 1

Supplemental Table 1. List of studies included in the independent participant data meta-analysis						
Study	Design	Psychiatric population	Primary outcome	Psilocybin arms	DD1 to endpoint	Sample size
Carhart-Harris et al., 2016	Open-label single-arm trial	TRD	QIDS-SR-16	10 mg and 25 mg, 7 days apart	Six weeks	19
Carhart-Harris et al., 2021	Randomized, placebo and active comparator-controlled trial	MDD	QIDS-SR-16	25 mg and 25 mg, 3 weeks apart	Six weeks	59
Davis et al., 2021	Randomized, waitlist-controlled trial	MDD	GRID-HAMD	20 mg/70 kg and 30 mg/70 kg, 1.6 weeks apart	Six weeks	24

Note: TRD = Treatment-Resistant Depression; MDD = Major Depressive Disorder; GRID-HAMD = GRID-Hamilton Depression Rating Scale; QIDS-SR-16 = Quick Inventory of Depression Symptoms Self-Report – 16; DD1 = Dosing Day 1. Note: the sample size describes the number of participants who enrolled in the study and completed the measures at both baseline and at six-week follow-up. The full sample from Carhart-Harris and colleagues (2016) is reported in Carhart-Harris and colleagues (2018).

Supplemental Table 2

Supplemental Table 2. Results of Cochrane risk of bias assessment						
Study	Selection bias		Performance bias	Detection bias	Attrition bias	Reporting bias
	Random sequence	Allocation concealment	Participant and personnel blinding	Outcome blinding	Incomplete outcome data	Selective reporting
Carhart-Harris et al., 2016	High	High	High	High	High	Low
Carhart-Harris et al., 2021	Low	Low	Unclear	Unclear	Low	Low
Davis et al., 2021	Low	Low	High	Low	High	Low

Note: Domains drawn from Cochrane guidelines (Higgins et al., 2011). The risk of bias assessment of Carhart-Harris and colleagues (2016) includes Carhart-Harris and colleagues (2018).

Supplemental Table 3

Supplemental Table 3. Demographic characteristics of included samples					
	Age	Male gender	White race/ethnicity	Undergraduate degree or higher	Unemployed
	M (SD)	n (%)	n (%)	n (%)	n (%)
All samples	41.55 (11.23)	59 (57.84%)	89 (87.25%)	79 (77.45%)	27 (26.47%)
Carhart-Harris et al., 2016	44.75 (10.88)	13 (68.42%)	15 (78.95%)	16 (84.21%)	10 (52.63%)
Carhart-Harris et al., 2021	41.22 (10.91)	39 (66.10%)	52 (88.14%)	44 (74.58%)	12 (20.34%)
Davis et al., 2021	39.83 (12.23)	7 (29.17%)	22 (91.67%)	19 (79.17%)	5 (20.83%)

Note: The descriptive statistics in this table only describe the participants who completed the measures at both baseline and at six-week follow-up. The raw data was recoded the categories that were deemed most appropriate. The full sample from Carhart-Harris and colleagues (2016) is reported in Carhart-Harris and colleagues (2018).

Supplemental Table 4

Supplemental Table 4. Descriptive statistics by study and treatment arm						
Study ID	Treatment arm	n	SMD Mean	SMD SD	# worse	% worse
Carhart et al., 2016	Psilocybin	19	-2.26	1.38	1	5.3
Carhart et al., 2021	Psilocybin	30	-1.88	1.51	4	13.3
Davis et al., 2021	Psilocybin	13	-3.70	1.91	1	7.7
Carhart et al., 2021	Escitalopram	29	-1.56	1.36	3	10.3
Davis et al., 2021	Waitlist	11	0.26	1.06	7	63.6

Note: Mean = average pre-post change in depression in standardized mean difference (SMD) units; SD = standard deviation of SMD change; # worse = number of participants in each treatment arm showing worsening symptoms ($SMD \geq 0.24$); % worse = percentage of participants in each treatment arm showing worsening symptoms. The full sample from Carhart-Harris and colleagues (2016) is reported in Carhart-Harris and colleagues (2018).

Supplemental Table 5

Supplemental Table 5. Predictors of response to psilocybin				
Predictor	B	B _{lb}	B _{ub}	p

Age	0.01	-0.03	0.04	.615
Male gender	0.20	-0.60	1.09	.638
White race/ethnicity	0.27	-1.15	1.61	.703
Undergraduate degree or higher	0.07	-0.91	1.01	.884
Unemployed	-0.08	-0.95	0.80	.853
Note: B = unstandardized regression coefficient; lb = lower bound of 95% confidence interval; ub = upper bound of 95% confidence interval; p = p-value.				

Supplemental Table 6

Supplemental Table 6. Predictors of symptom worsening				
Predictor	OR	OR _{lb}	OR _{ub}	p
Age	1.02	0.95	1.11	.525
Male gender	4.03	0.60	79.87	.216
White race/ethnicity	1.54	0.13	214.53	.770
Undergraduate degree or higher	0.24	0.04	1.47	.111
Unemployed	0.39	0.02	2.65	.405
Note: OR = odds ratio; lb = lower bound of 95% confidence interval; ub = upper bound of 95% confidence interval; p = p-value.				

Supplemental Materials. Deviations from preregistration

We made the following three deviations from our preregistration. As data were only available from the three studies focused on depression, we only examined changes in depression symptoms. As data from the six-week follow-up point were available from all included studies, we focused on this time point. As none of the included studies shared a control condition type, we did not examine moderation of treatment effects by baseline demographics.

Supplemental Materials. Data extraction

After removal of duplicates, the abstracts, titles, and/or keywords were independently screened by two authors (OS and PC) using the reported inclusion and exclusion criteria. The full texts of identified articles were subsequently screened based on the same eligibility criteria. Discrepancies were discussed between OS and PC until a consensus was reached. The corresponding author of each eligible study was contacted, and standardized measures of depression were requested.

Supplemental Materials. Additional information about included trials

All three studies administered two doses of psilocybin at different time points and with varying dosages. In Carhart-Harris and colleagues (2016), dosing day 1 (10 mg) and dosing day 2 (25 mg) were 7 days apart. In Carhart-Harris and colleagues (2021), dosing day 1 (25 mg) and dosing day 2 (25 mg) were 3 weeks apart. In Davis and colleagues (2021), there was a mean of 1.6 weeks between dosing day 1 (20mg/70kg) and dosing day 2 (30mg/70kg). All three studies assessed participants approximately 6 weeks post-dosing day 1, which was the chosen endpoint for this study.

Supplemental Materials. R code for analyses

```
#Psilocybin IDPMA
library(lm.beta)
library(lme4)
library(lmerTest)
library(psych)

#Reading in data
df <- read.csv("~/IPDMA Excel Data_Final.csv")

#recoding variables
df$Condition <- relevel(factor(df$Condition), ref="P")

table(df$Ethnicity)

df$White <- ifelse(df$Ethnicity=="White", yes=1, no=0)
```

```

table(df$White)

table(df$Employment.recoded)
df$Unemployed <- ifelse(df$Employment.recoded=="Unemployed",yes=1,no=0)
table(df$Unemployed)

#calculate effect sizes
df$QIDS_Baseline_SD <- sd(df$QIDS_Baseline,na.rm=TRUE)
df$QIDS_6weeks_change <- df$QIDS_6weeks - df$QIDS_Baseline
df$QIDS_6weeks_d <- df$QIDS_6weeks_change/df$QIDS_Baseline_SD

df$HAMD_Baseline_SD <- sd(df$HAMD_Baseline,na.rm=TRUE)
df$HAMD_5weeks_change <- df$HAMD_5weeks - df$HAMD_Baseline
df$HAMD_5weeks_d <- df$HAMD_5weeks_change/df$HAMD_Baseline_SD

df$HAMD_8weeks_change <- df$HAMD_8weeks - df$HAMD_Baseline
df$HAMD_8weeks_d <- df$HAMD_8weeks_change/df$HAMD_Baseline_SD

df$post_d_orig <- ifelse(!is.na(df$QIDS_6weeks_d),yes=df$QIDS_6weeks_d,no=
ifelse(!is.na(df$HAMD_5weeks_d),yes=df$HAMD_5weeks_d,no=NA))

df$post_d <- ifelse(!is.na(df$QIDS_6weeks_d),yes=df$QIDS_6weeks_d,no=
ifelse(!is.na(df$HAMD_8weeks_d),yes=df$HAMD_8weeks_d,no=NA))

#calculate worsening
df$post_d_worse <- ifelse(df$post_d>=.24,yes=1,no=0)
df$post_d_worse_orig <- ifelse(df$post_d_orig>=.24,yes=1,no=0)

table(df[,c("post_d_worse","post_d_worse_orig")])

table(df[,c("post_d_worse","post_d_worse_orig","Study.ID")])

#study ID recoding

```

```

df$btwn.grp <- ifelse(!df$Study.ID=="Carhart et al., 2016",yes=1,no=0)
table(df[,c("Study.ID","btwn.grp")])

#Descriptives
psych::describeBy(df[,c("post_d")],df$Condition)
prop.table(table(df[,c("post_d_worse","Condition")]),margin=2)

#Predictors of treatment response####
vars <- c("Age","Gender","White","Education.recoded","Unemployed")
tab.post.d <- data.frame(predictor = vars, est = NA, lb.ci = NA, ub.ci =
NA, se = NA, df = NA, t = NA, p = NA)

summary(lmer(post_d ~ Age + (1|Study.ID), data = df[df$Condition=="P",]))
#ns

summary(lmer(post_d ~ Gender + (1|Study.ID), data =
df[df$Condition=="P",])) #ns

summary(lmer(post_d ~ White + (1|Study.ID), data = df[df$Condition=="P",]))
#ns

summary(lmer(post_d ~ Education.recoded + (1|Study.ID), data =
df[df$Condition=="P",])) #ns

summary(lmer(post_d ~ Unemployed + (1|Study.ID), data =
df[df$Condition=="P",])) #ns

out.Age <- lmer(post_d ~ Age + (1|Study.ID), data = df[df$Condition=="P",])
#ns

out.Gender <- lmer(post_d ~ Gender + (1|Study.ID), data =
df[df$Condition=="P",]) #ns

out.White <- lmer(post_d ~ White + (1|Study.ID), data =
df[df$Condition=="P",]) #ns

out.Education.recoded <- lmer(post_d ~ Education.recoded + (1|Study.ID),
data = df[df$Condition=="P",]) #ns

out.Unemployed <- lmer(post_d ~ Unemployed + (1|Study.ID), data =
df[df$Condition=="P",]) #ns

tab.post.d[tab.post.d$predictor=="Age",c("est","se","df","t","p")] <-
summary(out.Age)$coefficients[2,]

tab.post.d[tab.post.d$predictor=="Gender",c("est","se","df","t","p")] <-
summary(out.Gender)$coefficients[2,]

tab.post.d[tab.post.d$predictor=="White",c("est","se","df","t","p")] <-
summary(out.White)$coefficients[2,]

```

```

tab.post.d[tab.post.d$predictor=="Education.recoded",c("est","se","df","t",
"p")] <- summary(out.Education.recoded)$coefficients[2,]

tab.post.d[tab.post.d$predictor=="Unemployed",c("est","se","df","t","p")]
<- summary(out.Unemployed)$coefficients[2,]

tab.post.d[tab.post.d$predictor=="Age",c("lb.ci","ub.ci")] <-
confint(out.Age)[4,]

tab.post.d[tab.post.d$predictor=="Gender",c("lb.ci","ub.ci")] <-
confint(out.Gender)[4,]

tab.post.d[tab.post.d$predictor=="White",c("lb.ci","ub.ci")] <-
confint(out.White)[4,]

tab.post.d[tab.post.d$predictor=="Education.recoded",c("lb.ci","ub.ci")] <-
confint(out.Education.recoded)[4,]

tab.post.d[tab.post.d$predictor=="Unemployed",c("lb.ci","ub.ci")] <-
confint(out.Unemployed)[4,]

tab.post.d[,c("est","lb.ci","ub.ci")] <-
round(tab.post.d[,c("est","lb.ci","ub.ci")],2)

tab.post.d$p <- round(tab.post.d$p,3)

tab.post.d

write.csv(tab.post.d[,c("est","lb.ci","ub.ci","p")], "~/Data/tab.post.d.csv"
, row.names = FALSE)

#Predictors of symptom worsening####
#psilocybin vs. control groups
table(df[,c("Study.ID","btwn.grp")])

summary(glmer(post_d_worse ~ Condition + (1|Study.ID), data =
df[df$btwn.grp==1,], family = "binomial"))

out <- glmer(post_d_worse ~ Condition + (1|Study.ID), data =
df[df$btwn.grp==1,], family = "binomial")

summary(out)

confint(out)

exp(out@beta[2]) #vs. escitalopram
exp(confint(out)[3,1])
exp(confint(out)[3,2])

exp(out@beta[3]) #vs. waitlist

```

```

exp(confint(out)[4,1])
exp(confint(out)[4,2])

summary(glmer(post_d_worse ~ Condition + (1|Study.ID) + Age + Gender +
White + Education.recoded + Unemployed,
            data = df[df$btwn.grp==1,], family = "binomial"))
out <- glmer(post_d_worse ~ Condition + (1|Study.ID) + Age + Gender + White
+ Education.recoded + Unemployed,
            data = df[df$btwn.grp==1,], family = "binomial")
summary(out)
confint(out)

exp(out@beta[2]) #vs. escitalopram
exp(confint(out)[3,1])
exp(confint(out)[3,2])

exp(out@beta[3]) #vs. waitlist
exp(confint(out)[4,1])
exp(confint(out)[4,2])

#examine cell sizes with dichotomous predictors
table(df[df$Condition=="P",c("post_d_worse","Gender")]) #previously empty
cell
table(df[df$Condition=="P",c("post_d_worse","White")]) #empty cell
table(df[df$Condition=="P",c("post_d_worse","Education.recoded")])
table(df[df$Condition=="P",c("post_d_worse","Unemployed")])

#demographic predictors of post_d
tab.post.d.worse <- data.frame(predictor = vars, or = NA, or.lb.ci = NA,
or.ub.ci = NA, est = NA, lb.ci = NA, ub.ci = NA, se = NA, z = NA, p = NA)

summary(glmer(post_d_worse ~ Age + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial")) #ns
summary(glmer(post_d_worse ~ Gender + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial")) #ns
summary(glmer(post_d_worse ~ White + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial")) #note SE

```

```

summary(glmer(post_d_worse ~ Education.recoded + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial"))

summary(glmer(post_d_worse ~ Unemployed + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial"))

out.Age <- glmer(post_d_worse ~ Age + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial") #ns

out.Gender <- glmer(post_d_worse ~ Gender + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial") #ns

out.White <- glmer(post_d_worse ~ White + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial") #ns

out.Education.recoded <- glmer(post_d_worse ~ Education.recoded +
(1|Study.ID), data = df[df$Condition=="P",], family = "binomial") #ns

out.Unemployed <- glmer(post_d_worse ~ Unemployed + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial") #ns

tab.post.d.worse[tab.post.d.worse$predictor=="Age",c("est","se","z","p")]
<- summary(out.Age)$coefficients[2,]

tab.post.d.worse[tab.post.d.worse$predictor=="Gender",c("est","se","z","p")
] <- summary(out.Gender)$coefficients[2,]

#tab.post.d.worse[tab.post.d.worse$predictor=="White",c("est","se","z","p")
] <- summary(out.White)$coefficients[2,]

tab.post.d.worse[tab.post.d.worse$predictor=="Education.recoded",c("est","s
e","z","p")] <- summary(out.Education.recoded)$coefficients[2,]

tab.post.d.worse[tab.post.d.worse$predictor=="Unemployed",c("est","se","z",
"p")] <- summary(out.Unemployed)$coefficients[2,]

tab.post.d.worse[tab.post.d.worse$predictor=="Age",c("or")] <-
exp(summary(out.Age)$coefficients[2,1])

tab.post.d.worse[tab.post.d.worse$predictor=="Gender",c("or")] <-
exp(summary(out.Gender)$coefficients[2,1])

#tab.post.d.worse[tab.post.d.worse$predictor=="White",c("or")] <-
exp(summary(out.White)$coefficients[2,1])

tab.post.d.worse[tab.post.d.worse$predictor=="Education.recoded",c("or")]
<- exp(summary(out.Education.recoded)$coefficients[2,1])

tab.post.d.worse[tab.post.d.worse$predictor=="Unemployed",c("or")] <-
exp(summary(out.Unemployed)$coefficients[2,1])

tab.post.d.worse[tab.post.d.worse$predictor=="Age",c("lb.ci","ub.ci")] <-
confint(out.Age)[3,]

tab.post.d.worse[tab.post.d.worse$predictor=="Gender",c("lb.ci","ub.ci")]
<- confint(out.Gender)[3,]

```

```

#tab.post.d.worse[tab.post.d.worse$predictor=="White",c("lb.ci","ub.ci")]
<- confint(out.White)[3,]

tab.post.d.worse[tab.post.d.worse$predictor=="Education.recoded",c("lb.ci",
"ub.ci")] <- confint(out.Education.recoded)[3,]

tab.post.d.worse[tab.post.d.worse$predictor=="Unemployed",c("lb.ci","ub.ci"
)] <- confint(out.Unemployed)[3,]

tab.post.d.worse$or.lb.ci <- exp(tab.post.d.worse$lb.ci)
tab.post.d.worse$or.ub.ci <- exp(tab.post.d.worse$sub.ci)

tab.post.d.worse

#Firth's method of penalized logistic regression for zero cells
library(logistf)

out.White <- logistf(post_d_worse ~ White + Study.ID, data =
df[df$Condition=="P",])

summary(out.White)

tab.post.d.worse[tab.post.d.worse$predictor=="White","est"] <-
out.White$coefficients[2]

tab.post.d.worse[tab.post.d.worse$predictor=="White","or"] <-
exp(out.White$coefficients[2])

tab.post.d.worse[tab.post.d.worse$predictor=="White","se"] <-
sqrt(out.White$var[2,2])

tab.post.d.worse[tab.post.d.worse$predictor=="White","z"] <- 0.09 #need to
hand code

tab.post.d.worse[tab.post.d.worse$predictor=="White","p"] <-
out.White$prob[2]

tab.post.d.worse[tab.post.d.worse$predictor=="White","or.lb.ci"] <-
exp(confint(out.White)[2,1])

tab.post.d.worse[tab.post.d.worse$predictor=="White","or.ub.ci"] <-
exp(confint(out.White)[2,2])

tab.post.d.worse[,c("or","or.lb.ci","or.ub.ci")] <-
round(tab.post.d.worse[,c("or","or.lb.ci","or.ub.ci")],2)

tab.post.d.worse$p <- round(tab.post.d.worse$p,3)

tab.post.d.worse

```



```
write.csv(tab.post.d.worse[,c("or","or.lb.ci","or.ub.ci","p")], "~/Data/tab.
post.d.worse.csv", row.names = FALSE)
```

```
#generating histogram of pre-post ds
```

```
names(df)
```

```
table(df$Condition)
```

```
df$ConditionR <-
```

```
car::recode(df$Condition, "'P'='Psilocybin'; 'E'='Escitalopram'; 'W'='Waitlist
'")
```

```
df$ConditionR <- relevel(factor(df$ConditionR), ref="Psilocybin")
```

```
ggplot(df, aes(x = post_d)) +
```

```
  geom_histogram() + theme_bw() +
```

```
  geom_vline(xintercept = 0.23, linetype="dashed") +
```

```
  facet_wrap(~ConditionR, ncol = 1) + ylab("Count") +
```

```
  xlab("Standardized Mean Difference")
```

```
ggsave("~/Data/psilo_ipd.pdf", height = 5.8, width = 3.3)
```

```
#comparing TRD to MDD psilocybin response
```

```
table(df$Study.ID)
```

```
df$trd <- ifelse(df$Study.ID=="Carhart et al., 2016", yes=1, no=0)
```

```
summary(glmmer(post_d_worse ~ trd + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial"))
```

```
out <- glmmer(post_d_worse ~ trd + (1|Study.ID), data =
df[df$Condition=="P",], family = "binomial")
```

```
summary(out)
```

```
confint(out)
```

```
exp(out@beta[2]) #vs. non-trd
```

```
exp(confint(out)[3,1])
```

```
exp(confint(out)[3,2])
```

```
table(df[,c("trd", "post_d_worse")])
```

```
prop.table(table(df[,c("trd", "post_d_worse")]), margin=1)
```

```
#additional descriptives
```

```
describeBy(df$post_d, df[,c("Study.ID", "Condition")])
```

```

out <- describeBy(df[df$Condition=="P", "post_d"], group =
df[df$Condition=="P", c("Study.ID")])

out <- data.frame(do.call("rbind", out))

out <- round(out, 2)

out$var <- rownames(out)

out <- out[, c("var", "n", "mean", "sd", "se")]

out$Condition <- "P"

out

out.P <- out

out <- describeBy(df[df$Condition=="E", "post_d"], group =
df[df$Condition=="E", c("Study.ID")])

out <- data.frame(do.call("rbind", out))

out <- round(out, 2)

out$var <- rownames(out)

out <- out[, c("var", "n", "mean", "sd", "se")]

out$Condition <- "E"

out

out.E <- out

out <- describeBy(df[df$Condition=="W", "post_d"], group =
df[df$Condition=="W", c("Study.ID")])

out <- data.frame(do.call("rbind", out))

out <- round(out, 2)

out$var <- rownames(out)

out <- out[, c("var", "n", "mean", "sd", "se")]

out$Condition <- "W"

out

out.W <- out

out.all <- rbind(out.P, out.E, out.W)

table(df[, c("Study.ID", "post_d_worse", "Condition")])

prop.table(table(df[, c("Study.ID", "post_d_worse", "Condition")]), margin=1)

```

```

out.all$worse_n <- NA
out.all$worse_per <- NA

#psilocybin

out.all[out.all$var=="Carhart et al., 2016" &
out.all$Condition=="P","worse_n"] <-

  table(df[df$Study.ID=="Carhart et al., 2016" &
df$Condition=="P","post_d_worse"])[2]

out.all[out.all$var=="Carhart et al., 2021" &
out.all$Condition=="P","worse_n"] <-

  table(df[df$Study.ID=="Carhart et al., 2021" &
df$Condition=="P","post_d_worse"])[2]

out.all[out.all$var=="Davis et al., 2021" &
out.all$Condition=="P","worse_n"] <-

  table(df[df$Study.ID=="Davis et al., 2021" &
df$Condition=="P","post_d_worse"])[2]

out.all[out.all$var=="Carhart et al., 2016" &
out.all$Condition=="P","worse_per"] <-

  prop.table(table(df[df$Study.ID=="Carhart et al., 2016" &
df$Condition=="P","post_d_worse"])[2])

out.all[out.all$var=="Carhart et al., 2021" &
out.all$Condition=="P","worse_per"] <-

  prop.table(table(df[df$Study.ID=="Carhart et al., 2021" &
df$Condition=="P","post_d_worse"])[2])

out.all[out.all$var=="Davis et al., 2021" &
out.all$Condition=="P","worse_per"] <-

  prop.table(table(df[df$Study.ID=="Davis et al., 2021" &
df$Condition=="P","post_d_worse"])[2])

#escitalopram

out.all[out.all$var=="Carhart et al., 2021" &
out.all$Condition=="E","worse_n"] <-

  table(df[df$Study.ID=="Carhart et al., 2021" &
df$Condition=="E","post_d_worse"])[2]

out.all[out.all$var=="Carhart et al., 2021" &
out.all$Condition=="E","worse_per"] <-

  prop.table(table(df[df$Study.ID=="Carhart et al., 2021" &
df$Condition=="E","post_d_worse"])[2])

```

```
#waitlist

out.all[out.all$var=="Davis et al., 2021" &
out.all$Condition=="W","worse_n"] <-

  table(df[df$Study.ID=="Davis et al., 2021" &
df$Condition=="W","post_d_worse"])[2]

out.all[out.all$var=="Davis et al., 2021" &
out.all$Condition=="W","worse_per"] <-

  prop.table(table(df[df$Study.ID=="Davis et al., 2021" &
df$Condition=="W","post_d_worse"]))[2]

out.all

out.all$worse_per <- round(out.all$worse_per*100,1)

write.csv(out.all[,c("var","Condition","n","mean","sd","worse_n","worse_per
")], "~/Data/out_means.csv")
```