Original Paper

Efficacy and safety of psilocybin-assisted treatment for major depressive disorder: Prospective 12-month follow-up

Journal of Psychopharmacology
2022, Vol. 36(2) 151–158

DOI: 10.1177/02698811211073759
journals.sagepub.com/home/jop
SAGE

Natalie Gukasyan[1], Alan K Davis[1,2], Frederick S Barrett[1], Mary P Cosimano[1], Nathan D Sepeda[1], Matthew W Johnson[1] and Roland R Griffiths[1,3]

Abstract

Background: Preliminary data suggest that psilocybin-assisted treatment produces substantial and rapid antidepressant effects in patients with major depressive disorder (MDD), but little is known about long-term outcomes.

Aims: This study sought to examine the efficacy and safety of psilocybin through 12 months in participants with moderate to severe MDD who received psilocybin.

Methods: This randomized, waiting-list controlled study enrolled 27 patients aged 21–75 with moderate to severe unipolar depression (GRID-Hamilton Depression Rating Scale (GRID-HAMD) ⩾ 17). Participants were randomized to an immediate or delayed (8 weeks) treatment condition in which they received two doses of psilocybin with supportive psychotherapy. Twenty-four participants completed both psilocybin sessions and were followed through 12 months following their second dose.

Results: All 24 participants attended all follow-up visits through the 12-month timepoint. Large decreases from baseline in GRID-HAMD scores were observed at 1-, 3-, 6-, and 12-month follow-up (Cohen $d = 2.3$, 2.0, 2.6, and 2.4, respectively). Treatment response (⩾50% reduction in GRID-HAMD score from baseline) and remission were 75% and 58%, respectively, at 12 months. There were no serious adverse events judged to be related to psilocybin in the long-term follow-up period, and no participants reported psilocybin use outside of the context of the study. Participant ratings of personal meaning, spiritual experience, and mystical experience after sessions predicted increased well-being at 12 months, but did not predict improvement in depression.

Conclusions: These findings demonstrate that the substantial antidepressant effects of psilocybin-assisted therapy may be durable at least through 12 months following acute intervention in some patients.

Keywords

Insight, long-term effects, major depressive disorder, mystical experience, psilocybin

Major depressive disorder (MDD) affects over 260 million people worldwide and is a leading cause of disability and healthcare expenditures (James et al., 2018). First-line treatments, including pharmacotherapy and psychotherapy, may take weeks or months to produce clinically meaningful symptom reduction, and patients can have difficulty with treatment adherence (Cuijpers et al., 2008; Kolovos et al., 2017; Lam, 2012). At least 30% of patients ultimately meet criteria for treatment-resistant depressive illness after failing to respond to multiple attempts at treatment (Nemeroff, 2007). MDD also has a highly recurrent course, with 40–60% of those diagnosed with a single episode eventually relapsing, and rate of relapse increasing with each subsequent episode (Richards, 2011; Solomon et al., 2000). Novel interventions are needed that can act rapidly and produce sustained remission.

Several preliminary studies suggest that psilocybin-assisted treatment may have substantial antidepressant effects in patients with MDD, with treatment response occurring within a week of administration of just one or two doses in the context of psychotherapy (Carhart-Harris et al., 2018, 2021; Davis et al., 2021). In an initial report of primary outcomes following two doses of psilocybin using a randomized waitlist-control study design, we reported a large effect size (Cohen $d = 2.3$) and high rates of treatment response and remission (71% and 54%) at 1 month following intervention (Davis et al., 2021). Treatment-resistant patients also appear to have a favorable response rate (Carhart-Harris et al., 2018). A more recent study used a double-blind

[1]Center for Psychedelic and Consciousness Research, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[2]College of Social Work, The Ohio State University, Columbus, OH, USA
[3]Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Corresponding authors:
Natalie Gukasyan, Center for Psychedelic and Consciousness Research, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, 5510 Nathan Shock Drive, Baltimore, MD 21224, USA.
Email: gukasyan@jhmi.edu

Roland R Griffiths, Center for Psychedelic and Consciousness Research, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, 5510 Nathan Shock Drive, Baltimore, MD 21224, USA.
Email: rgriff@jhmi.edu

double dummy design to compare high-dose psilocybin plus 6 weeks of placebo with very low dose psilocybin plus 6 weeks of escitalopram (Carhart-Harris et al., 2021). The authors failed to show a significant difference between the two groups at 6 weeks in their designated primary outcome measure (Quick Inventory of Depressive Symptoms). The majority of results for secondary outcome measures including other depression severity scores favored the high-dose psilocybin group, though analyses were not corrected for multiple comparisons.

Although psilocybin treatment of MDD appears promising, little is known about long-term efficacy and safety. The three studies conducted to date demonstrated efficacy at their longest follow-up assessments of 4 weeks (Davis et al., 2021), 6 weeks (Carhart-Harris et al., 2021), and 6 months (Carhart-Harris et al., 2018), although depression severity scores were trending upward at 3- and 6-month follow-up timepoints. Carhart-Harris et al. (2018), with the longest period of follow-up, had an open-label design. Given the chronicity and relapsing disease course of MDD (Richards, 2011), the present study represents a significant extension of these previous findings by assessing efficacy and safety of a psilocybin intervention throughout a 12-month follow-up period.

Methods

Study design

Full details of the study design and inclusion criteria have been described previously (Davis et al., 2021). All procedures involving subjects were approved by The Johns Hopkins Medicine Institutional Review Board. Written informed consent was obtained from all participants. Participants were 21–75 years of age, medically stable, and met criteria for a moderate to severe episode of MDD as defined by a score of ⩾17 on the GRID-Hamilton Depression Rating Scale (GRID-HAMD) assessed by blinded clinician raters. Individuals with personal or first- or second-degree relative history of psychotic or bipolar I or II disorder were excluded. To avoid interactions with psychoactive drugs including those used to treat depression, participants were required to refrain from using such medications for at least five half-lives before screening and for at least 1 month following the second psilocybin session. Following medical and psychological screening and baseline assessments, participants were randomized to an immediate or delayed treatment condition. Participants in the immediate treatment group began the intervention after screening, while those in the delayed treatment group began the intervention after an 8-week delay interval.

After participants entered the intervention period, they were provided with 6–8 h of preparatory meetings with two facilitators. At least one facilitator in each dyad had a master's or doctoral level of clinical training in mental health (e.g. master of social work, PhD in clinical psychology, MD specializing in psychiatry). Following preparation, participants received two doses of psilocybin at 20 mg/70 kg and 30 mg/70 kg spaced approximately 2 weeks apart. Psilocybin was administered in a comfortable room under the supervision of both facilitators following established safety guidelines (Johnson et al., 2008). A nondirective psychotherapeutic approach was taken on session days. Participants returned for follow-up at 1 day and 1 week following each drug administration session, and then at 1, 3, 6, and 12 months following the second session, during which depression severity was assessed with participant- and clinician-rated measures. Each follow-up visit included a 1–2 h meeting with at least one of the therapist facilitators. Functional magnetic resonance imaging was completed at baseline and 1 week after the second psilocybin session (Doss et al., 2021).

Outcome measures

Measures of depression severity. The primary outcome measure was the GRID-HAMD (Depression Rating Scale Standardization Team, 2003), which was assessed by blinded clinician raters via telephone as described previously (Davis et al., 2021). Inter-rater reliability at the 3-, 6-, and 12-month timepoints was 87.5% (see online supplement for additional information). Depression was also assessed with two self-report questionnaires: the Quick Inventory of Depressive Symptoms (QIDS) (Rush et al., 2003), and the Beck Depression Inventory II (BDI-II) (Beck et al., 1996). Depression severity was assessed at baseline and at each of the follow-up timepoints.

Participant-rated measures of acute psilocybin effects. Various measures of acute psilocybin effects assessed at the end of the session or the following day were reported previously (Davis et al., 2021). Of interest in this follow-up analysis was whether a subset of these acute measures would predict subsequent follow-up results. Based on previous studies showing associations between acute psilocybin measures and subsequent positive effects in healthy and patient samples (Bogenschutz et al., 2015; Davis et al., 2020; Garcia-Romeu et al., 2014; Griffiths et al., 2008, 2016, 2018), the following measures were examined: the Mystical Experience Questionnaire (MEQ30) (Barrett et al., 2015) and four single-item measures (Carbonaro et al., 2020) on which participants rated the degree to which the session experience was personally meaningful, spiritually significant, psychologically insightful, and psychologically challenging on a scale from 1 = no more than routine, everyday experiences to 8 = the single most (meaningful, spiritually significant, psychologically insightful, or psychologically challenging experience) of my life. The psychological challenge item was included as a comparison because it was not expected to predict subsequent positive outcomes. The MEQ30 was completed at the conclusion of each psilocybin session and the single-item measures were completed on the day following each session.

Overall well-being attributed to psilocybin. At the 1-, 3-, 6-, and 12-month follow-up timepoints, participants completed the Persisting Effects Questionnaire (Griffiths et al., 2018), which involved rating on a 6-point scale current persisting effects that they attributed to their psilocybin experiences (see online supplement for more information). For this study, an overall well-being score was calculated as the grand mean of the five subscales of positive change: attitudes about life, attitudes about self, mood, relationships, and behavior, with each expressed as a percentage of maximum possible score.

Safety measures. At each follow-up timepoint, adverse events were recorded, suicidal ideation was assessed using the Columbia Suicide Severity Rating Scale (C-SSRS) (Posner et al., 2008),

and symptoms indicative of hallucinogen persisting perceptual disorder (HPPD) were solicited (e.g. "Since your drug session have you experienced any uncontrolled or disturbing return of drug-like effects?").

Statistical analyses

A repeated-measures analysis of variance was conducted with time (baseline, 1 week post-treatment, and 1, 3, 6, and 12 months post-treatment) and condition (immediate and delayed treatment) as factors on the primary depression outcome (GRID-HAMD score), with effect sizes calculated using partial eta squared (η_p^2). This analysis showed a significant effect of time, but no significant effect of condition or a time-by-condition interaction. Therefore, data were collapsed across the conditions and a series of paired *t* tests compared baseline scores with scores at each of the follow-up timepoints with Bonferroni adjustment for multiple comparisons. Paired *t* test effect sizes were calculated using Cohen *d*. Descriptive statistics of follow-up measures were calculated, including treatment response (⩾50% reduction in depression scores from baseline) and remission (GRID-HAMD ⩽ 7, QIDS ⩽ 5, BDI-II ⩽ 9) for measures of depression (Beck et al., 1996; Rush et al., 2003; Zimmerman et al., 2013). The relationship between acute measures of session experiences and follow-up measures of acute psilocybin effects (MEQ30 and ratings of meaning, insight, spiritual significance and psychological challenge) were examined with Spearman's correlations (r_s). For these calculations, the highest ratings or scores from Session 1 and Session 2 for each participant were used. For the follow-up measures, the overall well-being score was expressed as a percentage of maximum possible score, and the depression measures were expressed as percentage change from the baseline score for each participant. Analyses of several group-based comparisons were conducted using χ^2 for categorical variables and *t* test for continuous variables. A two-tailed significance level of $p < .05$ was used for between-group comparisons and correlations. Analyses were completed using SPSS 26 and 27.

Results

Participants

As described in more detail previously (Davis et al., 2021), 27 participants were randomized and 24 completed both psilocybin sessions, with 13 and 11 assigned to the immediate and delayed treatment groups, respectively. All 24 participants completed all long-term follow-up assessment visits (see online supplement, CONSORT diagram). The group was 67% female and 92% Caucasian. One participant identified as Black and another as Asian; none identified as Hispanic. Participants had a mean (SD) age of 39.8 (12.2) years. Mean duration of illness (years since diagnosis of MDD) was 21.5 (12.2) years, and mean time in current major depressive episode was 24.4 (22.0) months. Of the participants, 88% had previously attempted treatment with an antidepressant (e.g. a selective serotonin, norepinephrine, or dopamine reuptake inhibitor, etc.) and 58% reported previous use of such medication in the current depressive episode. Twenty-five percent had previously used a psychedelic drug, with an average of 3.3 previous uses and an average time of 9.2 years since last use.

Changes in depressive symptoms

As reported previously (Davis et al., 2021), GRID-HAMD scores were significantly lower in the immediate treatment group at 1 and 4 weeks post treatment when compared with corresponding timepoints after randomization in the delayed treatment group. After completing the delay period, participants in the delayed treatment group completed the psilocybin intervention and follow-up assessments. In this follow-up study, analysis of variance with GRID-HAMD scores showed a significant effect of time (baseline and 5 post-treatment timepoints) ($F(4.4, 96.3) = 34.9$, $p < 0.001$; $\eta_p^2 = .61$), but no significant effect of condition (immediate vs delayed treatment) or a time-by-condition interaction. Thus, the following results are from data collapsed across conditions.

As shown in Figure 1, mean GRID-HAMD scores for the overall treatment sample decreased from a mean (SD) of 22.8 (3.9) at pretreatment baseline to 8.7 (7.6) at 1 week, 8.9 (7.4) at 4 weeks, 9.3 (8.8) at 3 months, 7.0 (7.7) at 6 months, and 7.7 (7.9) at 12 months post-treatment ($p > .001$ at all timepoints, paired *t* tests with Bonferroni correction). The effect sizes for these differences were large, with Cohen *d* (95% CI) being 2.3 (1.5, 3.1) at 1 week, 2.3 (1.5, 3.1) at 4 weeks, 2.0 (1.3, 2.7) at 3 months, 2.6 (1.7, 3.4) at 6 months, and 2.4 (1.6, 3.2) at 12 months. Similar significant, large magnitude, and sustained decreases in depression from pretreatment across the five follow-up assessments occurred with the two patient-rated depression assessment questionnaires (QIDS and BDI-II, online supplement Table S1, Figures S1 and S2).

As previously reported (Davis et al., 2021), at 1 week after treatment, 17 of the 24 participants (71%) showed a clinical response rate on the GRID-HAMD (⩾50% reduction from pretreatment) and 14 (58%) were in remission (GRID-HAMD score ⩽ 7) (Zimmerman et al., 2013). As shown in Table 1, the response and remission rates were generally sustained through the 12-month follow-up assessment, with final response and remission rates of 75% and 58%, respectively. Table 1 shows similar or greater response and remission rates with the two patient-rated measures of depression (QIDS and BDI-II).

Figure 2 shows GRID-HAMD depression scores for each of the 24 study participants from pretreatment through the 12-month follow-up assessment. Most participants showed large decreases in their depression score at the first follow-up interval at 1-week post-treatment, consistent with the 71% response rate and 58% remission rate shown in Table 1. The figure shows that psilocybin did not exacerbate depression in any participant and that 3 of 24 participants (13%) did not meet criteria for a treatment response at any post-treatment timepoint.

Figure 2 also provides detailed information about participants who started or resumed daily use of an antidepressant medication for depression (i.e. a selective serotonin reuptake inhibitor, a serotonin-norepinephrine reuptake inhibitor, or a norepinephrine-dopamine reuptake inhibitor) after psilocybin treatment. Of the 24 participants, 0 (0%), 3 (12.5%), 5 (20.8%), 8 (33.3%), and 8 (33.3%), respectively, reported daily antidepressant use at the 4-week and 3-, 6-, and 12-month follow-up assessments. The 8 participants who started antidepressant treatment by the 12-month timepoint had higher baseline GRID-HAMD scores (mean 25.3 vs 21.5, $p = .02$) compared with those who did not report antidepressant use; however, they were not statistically different in age, sex, years with depression, duration of current depressive

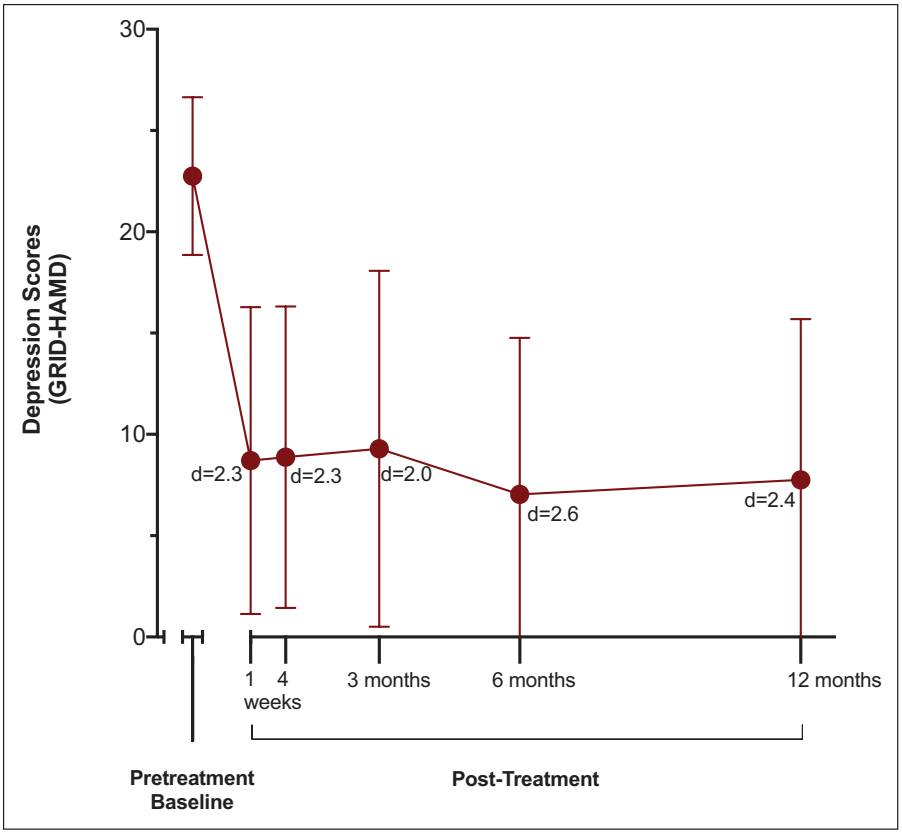


Figure 1. Decrease in GRID-HAMD depression scores over time from baseline through the 12-month follow-up ($N = 24$).
Data points are means and brackets are ±1 SD; lower brackets are truncated at GRID-HAMD scores of 0. Mean GRID-HAMD was 22.8 (3.9) at baseline, 8.7 (7.6) at 1 week, 8.9 (7.4) at 4 weeks, 9.3 (8.8) at 3 months, 7.0 (7.7) at 6 months, and 7.7 (7.9) at 12 months post-treatment. All timepoints were significantly different from baseline ($p < 0.001$). Cohen *d* effect size is shown for each timepoint. Cohen *d* (95% CI) was 2.3 (1.5–3.1) at 1 week, 2.3 (1.5–3.1) at 4 weeks, 2.0 (1.3–2.7) at 3 months, 2.6 (1.7–3.4) at 6 months, and 2.4 (1.6–3.2) at 12 months.

Table 1. Percentage of total sample ($N = 24$) meeting criteria for treatment response (reduction in depression ⩾ 50% from pretreatment baseline) or remission.

	Post-Treatment Follow-Up Time				
	1 week	4 weeks	3 months	6 months	12 months
GRID-HAMD					
Response rate	71%	71%	67%	79%	75%
Remission rate[a]	58%	54%	54%	71%	58%
QIDS					
Response rate	79%	71%	79%	79%	79%
Remission rate[b]	54%	54%	58%	67%	67%
BDI-II					
Response rate	79%	79%	79%	88%	83%
Remission rate[c]	67%	63%	58%	75%	75%

GRID-HAMD: Hamilton Depression Rating Scale; QIDS: Quick Inventory of Depressive Symptoms; BDI-II: Beck Depression Inventory II.
[a]Remission = GRID-HAMD score ⩽ 7 (Zimmerman et al., 2013).
[b]Remission = QIDS score ⩽ 5 (Rush et al., 2003).
[c]Remission = BDI-II score ⩽ 9 (Beck et al., 1996).

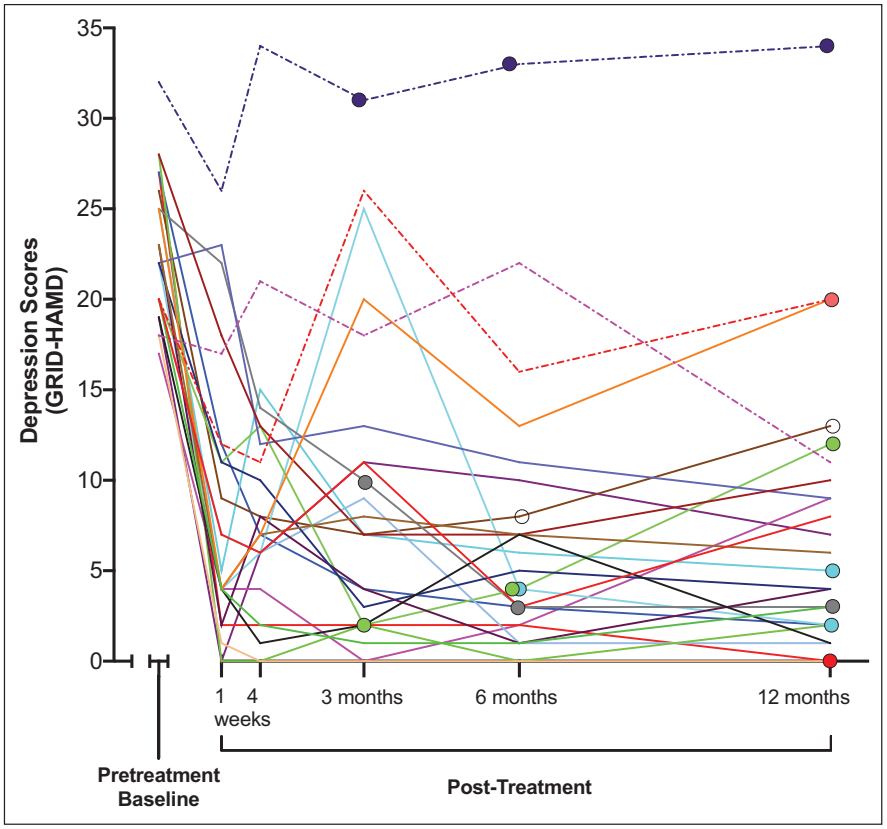


Figure 2. Depression scores (GRID-HAMD) for each of 24 study participants at baseline and each of 5 follow-up assessment timepoints. Individual participants are represented with different colors. Dashed lines indicate three participants who did not fulfill criteria for a treatment response at any post-treatment timepoint. Enlarged data points indicate participants who reported treatment with antidepressant medication, with the left-most enlarged data points showing the first follow-up timepoint at which medication use was reported.

episode, or history of medication use in the current depressive episode. Mean GRID-HAMD scores and overall well-being scores at 12 months did not significantly differ between those who did and did not begin antidepressant treatment.

Overall well-being attributed to psilocybin across the 1-, 3-, 6-, and 12-month follow-up timepoints was intermediate and stable. The grand mean (SD) overall well-being score, expressed as percentage of maximum possible score, was 63.9 (22.6), 60.0 (21.3), 59.0 (24.0), and 65.0 (20.0), respectively.

Participant-rated measures of session experiences as predictors of subsequent overall well-being and changes in depression severity

Correlations between participant-rated measures of psilocybin experiences at the time of the session and subsequent measures of well-being and depression were examined. At the first long-term follow-up assessment (Table 2, Week 4), ratings of personal meaning, psychological insight, spiritual significance, and the mystical experience (MEQ30) correlated significantly with well-being, and ratings of personal meaning and spiritual significance correlated significantly with improvement in depression (GRID-HAMD). However, at subsequent follow-up timepoints, none of these session experience measures were significantly correlated with improvements in depression (Table 2). The MEQ30 significantly correlated with well-being at all four follow-up timepoints, and ratings of personal meaning and spiritual significance were significantly correlated at three of four timepoints. Ratings of psychological challenge during the session were not significantly correlated with subsequent measures of well-being or depression at any follow-up timepoint.

Safety outcomes

During the follow-up period, there were no serious adverse events, suicidal ideation remained low, there were no instances of self-injurious behavior, no reported use of psilocybin or other psychedelics, and no participant met criteria for HPPD. Further details on adverse events are available in the online supplement.

Discussion

The present study suggests that two doses of psilocybin provided in the context of supportive therapy for MDD produced large and stable antidepressant effects throughout a 12-month follow-up

Table 2. Relationship between measures of psilocybin experiences assessed at the end of the session or the following day with the follow-up measures of well-being and improvement in depression assessed at 4 weeks and 3, 6, and 12 months.[a]

Session measures[b]	4 weeks	Follow-up	3 months	Follow-up	6 months	Follow-up	12 months	Follow-up
	Well-being[c]	GRID-HAMD[d]	Well-being[c]	GRID-HAMD[d]	Well-being[c]	GRID-HAMD[d]	Well-being[c]	GRID-HAMD[d]
Personal Meaning[e]	**0.70***	**0.67***	0.43	0.34	**0.51**	0.44	**0.45**	0.43
Psychological Insight[e]	**0.49**	0.36	0.35	0.04	0.27	0.34	0.39	0.25
Spiritual Significance[e]	**0.67***	**0.56***	**0.54**	0.16	0.44	0.28	**0.60***	0.40
Psychological Challenge[e]	0.12	0.32	0.09	0.18	0.13	0.31	0.06	0.13
Mystical Experience MEQ30	**0.71***	0.38	**0.43**	0.17	**0.50**	0.05	**0.50**	0.19

HAMD: Hamilton Depression Rating Scale; MEQ30: Mystical Experience Questionnaire.
[a]Data show Spearman's correlations (r_s); bold font indicates $p < .05$ and asterisks indicate $p < .01$.
[b]Assessments shown in this column were the highest rating or score from Sessions 1 and 2 for each participant.
[c]Well-being scores were expressed as percentage of maximum possible score.
[d]GRID-HAMD depression scores were expressed as percentage change from baseline for each participant.
[e]Only 20 of 24 participants completed these measures due to an error in the survey programming.

period. More specifically, depression, as measured by blinded clinician-rated assessments (GRID-HAMD), decreased substantially after treatment and remained low at 1, 3, 6, and 12 months post-treatment. The effect size at 12 months was very large (Cohen $d = 2.4$). Likewise, high and stable rates of response and remission occurred throughout the follow-up period (75% response and 58% remission at 12 months). Two patient-rated measures of depression (QIDS and BDI-II) showed similar large magnitude and stable antidepressant effects on mean scores and on response and remission rates. These findings suggesting enduring antidepressant effects of psilocybin 1 year after treatment significantly extend the previous results in this and two other trials that showed antidepressant effects through 4 weeks (Davis et al., 2021), 6 weeks (Carhart-Harris et al., 2021), and 6 months (Carhart-Harris et al., 2018). Notably, the remission rate and magnitude of the effect in the current study at both 6 and 12 months were substantially greater than those in a previous study at 6 months (QIDS, remission rate 67% and 67% vs 32%, Cohen *d* 2.2 and 2.3 vs 1.6, respectively) (Carhart-Harris et al., 2018). Whether this difference reflects population differences in severity of illness or procedural differences is unknown. Future research is needed to explore the possibility that efficacy of psilocybin treatment in MDD may be substantially longer than the 12 months observed in the present study, as has been suggested in a study that documented decreases in depressive symptoms up to 4.5 years following psilocybin treatment in patients with cancer-related distress (Agin-Liebes et al., 2020).

Of note, eight patients (33%) reported beginning a new course of treatment with a daily antidepressant drug at some point during the 12-month follow-up period, which is similar to the 32% of patients in a previous trial that did so by 6 months (Carhart-Harris et al., 2018). Although patients who used antidepressants during the follow-up period had higher GRID-HAMD scores at baseline, at 12 months they did not significantly differ from those who did not initiate medications. Determining the extent of the contribution of psilocybin vs other medications to clinical improvement in those who resumed antidepressant use is not possible. However, participant ratings of persisting well-being attributed to the psilocybin sessions were not significantly different from those who did not use antidepressant medications, suggesting that psilocybin treatment resulted in some independent benefit.

Although relapse and remission rates at 12 months were favorable, the ability to accurately compare the long-term efficacy of psilocybin-assisted treatment to that of standard antidepressant treatment is limited. The majority of recent studies of long-term antidepressant efficacy drop non-responders from follow-up and focus on rate of relapse among those who respond to a particular drug, which is often a minority of the intention-to-treat sample (McGrath et al., 2006; Trivedi et al., 2006). In our sample, of the 17 participants who met criteria for treatment response at 1-month follow-up, 12 (71%) continued to meet criteria for treatment response at all subsequent timepoints. An additional 3 participants who responded at 1 month also met criteria for treatment response at 12 months, but had one or more interim assessments during which their GRID-HAMD score was elevated out of the range for treatment response. The 71% continuous treatment response at 12 months is somewhat higher than the 54% rate reported in a study of fluoxetine responders who were maintained on fluoxetine, and much higher than those switched to placebo (28%) (McGrath et al., 2006).

The present study provides new information about qualitative features of the acute psilocybin experience that predict subsequent enduring effects. Patient ratings of personal meaning, spiritual significance, and MEQ30 scores after psilocybin sessions significantly correlated with a measure of overall well-being at most follow-up timepoints. However, except for ratings of personal meaning and spiritual significance at the first long-term follow-up assessment at 4 weeks, none of patient ratings of psilocybin experience at the time of the session were predictive of improvements in depression. Notably, two previous studies in individuals with cancer-related depression and anxiety (Griffiths et al., 2016; Ross et al., 2016) showed positive associations of MEQ30 session experiences with improvements in depression symptoms at 5 or 6 weeks. Considering that the direction of correlation at 4 weeks was in the predicted direction ($r_s = 0.38$, $p = .066$), it is possible that the present study was underpowered to detect such an effect with MEQ30 or other measures of psilocybin experience. Alternatively, this difference may reflect a lack of such relationship in a sample of individuals with MDD as opposed to depressive symptoms secondary to a cancer diagnosis.

There were no serious adverse events, depression symptoms were not significantly exacerbated in any participant, and there

was no reported use of psilocybin or other psychedelic drugs during the follow-up period. This latter observation contrasts with a previous study in which 5 of 19 participants reported use of psilocybin outside of the research setting by the end of the 6-month follow-up period (Carhart-Harris et al., 2018). Reasons for this difference are unknown, but the observation indicates the importance of assessing use of psychedelics outside of a clinical trial. Although the safety results presented herein are favorable, larger phase 3 and 4 studies will be needed to more fully assess safety.

Strengths and limitations

Strengths of this study of psilocybin-facilitated treatment of depression include a primary outcome measure that was assessed by blinded clinician raters, the longest post-treatment follow-up interval to date, and excellent participant retention. Although no participants reported extraneous psilocybin use, 33% reported using antidepressants during the follow-up period, which precludes determination of the effects of psilocybin alone in those patients. Although the randomized waiting list-control design of the study allowed for comparison of short-term treatment effects to the control group, as described previously (Davis et al., 2021), the design did not allow for a comparison group at long-term follow-up. A recent study suggests that expectancy effects and psychotherapy may account for some of the clinical benefit of psychedelic-assisted therapy (Carhart-Harris et al., 2021). In that study, which utilized a double-blind, double dummy design, both the high-dose and very low-dose psilocybin groups showed significant immediate decreases in depression, suggesting that the preparation and drug administration day procedures may reduce depressive symptoms even in the absence of high-dose psilocybin. Studies of other types of interventions for patients with MDD have demonstrated that placebo effects may last for weeks or months beyond intervention, and a lack of a comparator group makes it difficult to account for such effects in our study (Khan et al., 2008). Other limitations include the small sample size, the predominately Caucasian, non-Hispanic study sample, and exclusion of those judged to be at elevated risk of suicide.

Clinical implications

As novel antidepressants, classic psychedelics are commonly compared to ketamine and its analogues. Despite distinct mechanisms of pharmacologic action, both have rapid antidepressant effects and both have garnered concern about their potential for non-medical use (Schak et al., 2016; Shalit et al., 2019). Ketamine has nontrivial abuse potential and there may be overlap between mechanisms underlying its antidepressant effects and abuse potential, which may be exacerbated by the requirement for repeated administration to maintain therapeutic efficacy (Kokane et al., 2020; Liu et al., 2016). Although evidence to date suggests that psilocybin has relatively low abuse potential (Johnson et al., 2018), there remains concern for its potential to cause harm or encourage substance misuse in vulnerable populations (Reiff et al., 2020; Schatzberg, 2020). The present study highlights a key potential advantage of psilocybin treatment over ketamine in that antidepressant effects after just two administrations of psilocybin paired with psychological support appear to be sustained through 12 months, which is well beyond the duration of effects reported with ketamine (McIntyre et al., 2021; Salloum et al., 2020). It will be important for future research to determine the risks and benefits of additional psilocybin administration for those who failed to respond or experienced early relapse.

Conclusions

The results of this long-term follow-up of participants who were not blinded to the drug condition suggest that psilocybin-assisted treatment for MDD produces large and stable antidepressant effects throughout at least 12 months after treatment. These data document larger effects of longer duration than previous studies of psilocybin in depressed patients. Further studies are needed with active treatment or placebo comparison controls in larger and more diverse populations.

Acknowledgements

Darrick May, MD, Annie Umbricht, MD, and Eric Strain, MD, provided medical oversight during the study sessions. Patrick H. Finan, PhD, Jessiy Salwen, PhD, and Mary Bailes, LCPC, served as blinded clinician raters. Laura Doyle, BA, John Clifton, BS, Kasey Cox, MS, and Rhiannon Mayhugh, PhD, facilitated the intervention sessions and data collection. We thank Darrick G. May, MD, for help in the design and conduct of the study, and James B. Potash, MD, for his helpful comments on this article. These individuals, from Johns Hopkins University, received no additional compensation, outside of their usual salary, for their contributions.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: AKD is a board member of Source Research Foundation. MWJ has received grant support from the Heffter Research Institute unrelated to this study and he is an advisor to the following companies: AJNA Labs LLC, AWAKN Life Sciences Inc., Beckley Psytech Ltd., Entheon Biomedical Corp., Field Trip Psychedelics Inc., Mind Medicine Inc., Otsuka Pharmaceutical Development & Commercialization Inc., and Silo Pharma, Inc. RRG is a board member of the Heffter Research Institute and has received grant support from the Heffter Research Institute unrelated to this study. RRG is site principal investigator, and MWJ and NG are co-investigators for a multi-site trial of psilocybin-assisted therapy for major depressive disorder sponsored by Usona Institute.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported in part by a crowd-sourced funding campaign organized by Tim Ferriss and by grants from the Riverstyx Foundation and Dave Morin. Effort for AKD and NG was provided by NIH grant T32DA07209 from NIDA. Effort for authors was also provided by The Center for Psychedelic and Consciousness Research which is funded by the Steven and Alexandra Cohen Foundation, Tim Ferriss, Matt Mullenweg, Craig Nerenberg, and Blake Mycoskie. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ORCID iDs

Natalie Gukasyan https://orcid.org/0000-0003-3567-1421
Alan K Davis https://orcid.org/0000-0003-4770-8893
Roland R Griffiths https://orcid.org/0000-0001-5185-7854

Supplemental material

Supplemental material for this article is available online.

References

Agin-Liebes GI, Malone T, Yalch MM, et al. (2020) Long-term follow-up of psilocybin-assisted psychotherapy for psychiatric and existential distress in patients with life-threatening cancer. *Journal of Psychopharmacology* 34(2): 155–166.

Barrett FS, Johnson MW and Griffiths RR (2015) Validation of the revised Mystical Experience Questionnaire in experimental sessions with psilocybin. *Journal of Psychopharmacology* 29(11): 1182–1190.

Beck AT, Steer RA, Ball R, et al. (1996) Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *Journal of Personality Assessment* 67(3): 588–597.

Bogenschutz MP, Forcehimes AA, Pommy JA, et al. (2015) Psilocybin-assisted treatment for alcohol dependence: A proof-of-concept study. *Journal of Psychopharmacology* 29(3): 289–299.

Carbonaro TM, Johnson MW and Griffiths RR (2020) Subjective features of the psilocybin experience that may account for its self-administration by humans: A double-blind comparison of psilocybin and dextromethorphan. *Psychopharmacology* 237: 2293–2304.

Carhart-Harris RL, Bolstridge M, Day C, et al. (2018) Psilocybin with psychological support for treatment-resistant depression: Six-month follow-up. *Psychopharmacology* 235: 399–408.

Carhart-Harris RL, Giribaldi B, Watts R, et al. (2021) Trial of psilocybin versus escitalopram for depression. *New England Journal of Medicine* 384(15): 1402–1141.

Cuijpers P, Van Straten A, Andersson G, et al. (2008) Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology* 76(6): 909–922.

Davis AK, Barrett FS and Griffiths RR (2020) Psychological flexibility mediates the relations between acute psychedelic effects and subjective decreases in depression and anxiety. *Journal of Contextual Behavioral Science* 15: 39–45.

Davis AK, Barrett FS, May DG, et al. (2021) Effects of psilocybin-assisted therapy on major depressive disorder: A randomized clinical trial. *JAMA Psychiatry* 78(5): 481–489.

Depression Rating Scale Standardization Team (2003) *GRID-HAMD-17, GRID-HAMD-21 structured interview guide*. San Diego, CA: International Society for CNS Drug Development.

Doss MK, Považan M, Rosenberg MD, et al. (2021) Psilocybin therapy increases cognitive and neural flexibility in patients with major depressive disorder. *Translational Psychiatry* 11(1): 1–10.

Garcia-Romeu A, Griffiths RR and Johnson MW (2014) Psilocybin-occasioned mystical experiences in the treatment of tobacco addiction. *Current Drug Abuse Reviews* 7(3): 157–164.

Griffiths RR, Johnson MW, Carducci MA, et al. (2016) Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: A randomized double-blind trial. *Journal of Psychopharmacology* 30(12): 1181–1197.

Griffiths RR, Johnson MW, Richards WA, et al. (2018) Psilocybin-occasioned mystical-type experience in combination with meditation and other spiritual practices produces enduring positive changes in psychological functioning and in trait measures of prosocial attitudes and behaviors. *Journal of Psychopharmacology* 32(1): 49–69.

Griffiths RR, Richards WA, Johnson MW, et al. (2008) Mystical-type experiences occasioned by psilocybin mediate the attribution of personal meaning and spiritual significance 14 months later. *Journal of Psychopharmacology* 22(6): 621–632.

James SL, Abate D, Abate KH, et al. (2018) Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* 392(10159): 1789–1858.

Johnson MW, Griffiths RR, Hendricks PS, et al. (2018) The abuse potential of medical psilocybin according to the 8 factors of the Controlled Substances Act. *Neuropharmacology* 142: 143–166.

Johnson MW, Richards WA and Griffiths RR (2008) Human hallucinogen research: Guidelines for safety. *Journal of Psychopharmacology* 22(6): 603–620.

Khan A, Redding N and Brown WA (2008) The persistence of the placebo response in antidepressant clinical trials. *Journal of Psychiatric Research* 42(10): 791–796.

Kokane SS, Armant RJ, Bolaños-Guzmán CA, et al. (2020) Overlap in the neural circuitry and molecular mechanisms underlying ketamine abuse and its use as an antidepressant. *Behavioural Brain Research* 384: 112548.

Kolovos S, van Tulder MW, Cuijpers P, et al. (2017) The effect of treatment as usual on major depressive disorder: A meta-analysis. *Journal of Affective Disorders* 210: 72–81.

Lam RW (2012) Onset, time course and trajectories of improvement with antidepressants. *European Neuropsychopharmacology* 22: S492–S498.

Liu Y, Lin D, Wu B, et al. (2016) Ketamine abuse potential and use disorder. *Brain Research Bulletin* 126: 68–73.

McGrath PJ, Stewart JW, Quitkin FM, et al. (2006) Predictors of relapse in a prospective study of fluoxetine treatment of major depression. *American Journal of Psychiatry* 163(9): 1542–1548.

McIntyre RS, Rosenblat JD, Nemeroff CB, et al. (2021) Synthesizing the evidence for ketamine and esketamine in treatment-resistant depression: An international expert opinion on the available evidence and implementation. *American Journal of Psychiatry* 178(5): 383–399.

Nemeroff CB (2007) Prevalence and management of treatment-resistant depression. *Journal of Clinical Psychiatry* 68(8): 17–25.

Posner K, Brent D, Lucas C, et al. (2008) *Columbia-suicide severity rating scale (C-SSRS)*. New York: Columbia University Medical Center.

Reiff CM, Richman EE, Nemeroff CB, et al. (2020) Psychedelics and psychedelic-assisted psychotherapy. *American Journal of Psychiatry* 177(5): 391–410.

Richards D (2011) Prevalence and clinical course of depression: A review. *Clinical Psychology Review* 31(7): 1117–1125.

Ross S, Bossis A, Guss J, et al. (2016) Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: A randomized controlled trial. *Journal of Psychopharmacology* 30(12): 1165–1180.

Rush AJ, Trivedi MH, Ibrahim HM, et al. (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.

Salloum NC, Fava M, Hock RS, et al. (2020) Time to relapse after a single administration of intravenous ketamine augmentation in unipolar treatment-resistant depression. *Journal of Affective Disorders* 260: 131–139.

Schak KM, Vande Voort JL, Johnson EK, et al. (2016) Potential risks of poorly monitored ketamine use in depression treatment. *American Journal of Psychiatry* 173(3): 215–218.

Schatzberg AF (2020) Some comments on psychedelic research. *American Journal of Psychiatry* 177(5): 368–369.

Shalit N, Rehm J and Lev-Ran S (2019) Epidemiology of hallucinogen use in the US results from the National epidemiologic survey on alcohol and related conditions III. *Addictive Behaviors* 89: 35–43.

Solomon DA, Keller MB, Leon AC, et al. (2000) Multiple recurrences of major depressive disorder. *American Journal of Psychiatry* 157(2): 229–233.

Trivedi MH, Rush AJ, Wisniewski SR, et al. (2006) Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice. *American Journal of Psychiatry* 163(1): 28–40.

Zimmerman M, Martinez JH, Young D, et al. (2013) Severity classification on the Hamilton depression rating scale. *Journal of Affective Disorders* 150(2): 384–388.

Supplemental Material

Gukasyan N, Davis AK, Barrett FS, Cosimano MP, Sepeda ND, Johnson, MW, Griffiths, RR. Efficacy and safety of psilocybin-assisted treatment for major depressive disorder: Prospective twelve-month follow-up. *Journal of Psychopharmacology*

Supplemental Methods

Inter-rater Reliability for Primary Outcome Measure

Methods for assessing inter-rater reliability were identical to those described in Davis et al. (2021). Three clinician raters assessed the primary outcome measure for depression (GRID-HAMD) via telephone. All raters were employed at Johns Hopkins and did not conduct any other study assessments nor had any other study involvement. The three raters were trained on a set of standardized and practice interviews developed and conducted by an expert rater (author AKD). Raters were required to be within 3 points of the rating of the expert rater (author AKD) in order to proceed with conducting ratings in the trial. After each rater began rating depression in participants, each of their assessments were rated by another rater using audio recordings until they achieved three consecutively reliable ratings. A rating was considered reliable if the rating from the second rater was within 3 points of the rating by the primary rater. If a rating fell outside of this range, then the two raters met to review and discuss the audio recording and to mutually agree on a final rating. Following initial training and establishing reliability in assessment

measurement, ongoing inter-rater reliability was examined for each rater. This examination consisted of randomly selecting one assessment out of every ten assessments for each rater. A second rater listened to the audio recording of the selected assessment. If the rating from the second rater was within 3 points of the rating by the primary rater, the primary rating was used. If the rating was not within 3 points, then the two raters met to review the audio recording and to mutually agree on the final rating.

Assessment of Secondary Outcomes

In addition to the measures of depression severity described in the published report, follow-up data were obtained on the Patient Health Questionnaire (PHQ-9) (Kroenke et al., 2002). These results are not reported due to missing data. However, the results from analysis of the incomplete PHQ-9 data were concordant with the results reported from the other measures of depression severity.

As a safety measure, the Columbia Suicide Severity Rating Scale (CSSRS) (Posner et al., 2008, 2011) was administered at every in-person visit to assess for potentially worsening suicidality. The "Lifetime/Recent" version of the CSSRS was used at baseline and the "Since last visit" version for all subsequent administrations. Suicidal ideation (SI) severity is reported as the highest level of SI under the “suicidal ideation” assessment section of the CSSRS, which provides a 0-5 score, with higher score indicating higher severity as follows: 0=no passive or active SI; 1=passive SI; 2=nonspecific active SI; 3=active SI with any methods but no intent or plan; 4=Active SI with some intent, without specific plan; 5=active SI with specific plan and intent.

The Persisting Effects Questionnaire was a 154-item measure, expanded from prior versions used by our group (Griffiths et al., 2006, 2011) that assessed changes in attitudes, moods, behavior, and spiritual experience and has been shown to be to sensitive to the effects of psilocybin 14 months after a psilocybin session (Griffiths et al., 2011). Participants were instructed to rate any persisting effects that they considered were due to the experiences during their psilocybin session and their contemplation of those experiences. Items were rated on a six-point Likert scale (0=none, not at all; 1=so slight cannot decide; 2=slight; 3=moderate; 4=strong; 5=extreme, more than ever before in your life and stronger than 4). Within the questionnaire, the items were labeled in seven categories: Attitudes about life (13 positive and 13 negative items); Attitudes about self (11 positive and 11 negative items); Mood changes (9 positive and 9 negative items); Relationships (10 positive and 10 negative items); Behavioral changes (1 positive and 1 negative item); Spirituality (29 positive and 28 negative items); Miscellaneous (9 unscored items). The positive and negative items were intermixed within each category. For this study an overall well-being score was calculated as the grand mean of the 5 subscales of positive change: attitides about life, attitudes about self, mood, relationships, and behavior, with each expressed as a percentage of maximum possible score. A copy of the full questionnaire is available from the authors by request.

Supplemental Results and Discussion

CONSORT diagram

Figure S1 presents the CONSORT diagram of participant flow.

Secondary Depression Outcomes

Table S1 and eFigures 2 and 3 show results of primary (GRID-HAMD) and secondary (QIDS and BDI-II) depression measures for the entire treatment sample (N=24). Secondary measures of depression showed similar patterns of results to those of the GRID-HAMD with sustained decreases in symptoms throughout the follow-up period.

Safety Measures

Suicidal ideation: Complete data from the CSSRS was available for all participants except one whose 12-month assessment was missing. As previously reported (Davis et al., 2021), severity of suicidal ideation decreased significantly at 1 week after treatment and remained significantly reduced at 4 weeks (mean reduction of 1.0 points on CSSRS suicidal ideation question for both timepoints, paired sample *t* test $p<0.001$). CSSRS suicidal ideation score was also reduced from baseline at subsequent timepoints, though this difference was not significant at 12 months (see Table S1). There were no reported suicide attempts or instances of self-injurious behavior during the follow-up period.

Adverse events: Adverse events prior to the 2-week follow-up timepoint were previously reported (see Davis et al., 2021: Supplement 2, eTables 8 and 9). There were no serious adverse events recorded between the 2-week and 12-month follow-ups. Only one event was judged to be potentially related to study drug administration. A participant reported a one-minute long episode of visual distortions about three months following her second psilocybin session. She described an experience lasting about one minute, during which a patterned floor appeared to be “popping out in 3D.” The experience was described as pleasurable, did not meet criteria for HPPD, and did not recur. Her course was also notable for a higher than average number of recorded adverse events following psilocybin administration. These included post-session headache and residual psilocybin effects lasting into the day following her first session. These residual effects were described as increased intensity of color, heightened emotions, and an increased sense of equanimity. At one-week follow-up after her first session she also reported slight movement in her peripheral visual fields, altered perception of body sensations (feeling “disconnected” from her senses”), and vivid dreams. The participant’s medical history was notable for a viral respiratory illness about one month prior to psilocybin administration, during which she experienced tinnitus. Tinnitus has been identified as a possible risk factor for HPPD (Halpern et al., 2016). To our knowledge this is the first reported instance of a delayed perceptual distortion related to psychedelic use in a research setting in the modern era of clinical research with these drugs.

Other Secondary Outcomes

Participants rated high levels of persisting positive effects and very low levels of negative effects attributed to psilocybin session experiences assessed at 4 weeks and 3, 6, and 12 months on the Persisting Effects Questionnaire (data not shown). Mean scores on the subscales reflecting attitudes about life, attitudes about self, mood, relationships, behavior, and spirituality were quite stable with no apparent increasing or decreasing trend over time. Table S2 shows these data at

the 12 month follow-up, with mean rates of endorsing positive effects ranging from 51% to 75% of maximum possible score and mean rates of negative effects ranging from 0.8% to 6.7%. These rates of endorsing positive and negative effects are quite similar to those reported by cancer patients 6 months after psilocybin (Griffiths et al., 2016).

The relationship between measures of psilocybin experiences assessed at the end of the session or the following day with follow-up measures of well-being and depression was assessed to determine if acute measures at the time of the session were predictive of subsequent outcomes. The failure of a measure of challenging experience to be a strong predictor of well-being and therapeutic efficacy was expected based on prior observations (Carbonaro et al., 2016). As presented in Table 3 in the main manuscript, MEQ30 scores significantly correlated with well-being across all timepoints but did not significantly correlate with change in GRID-HAMD depression scores at any timepoint. The lack of significant correlation with depression outcome measures was unexpected because prior research in psychologically distressed cancer patients had shown such a relationship at five (Griffiths et al., 2016) or six (Ross et al., 2016) weeks post-treatment. Indeed, an earlier analysis of results in the present study (Davis et al., 2021: Supplement 2) had shown a moderate but significant correlation between MEQ30 scores and decreases in GRID-HAMD scores at 4 weeks when depression change scores in the delayed treatment group were expressed as a change from the immediate pre-session GRID-HAMD score rather than a change from the screening baseline score. When screening baseline scores are used to assess change in depression (Table 3), as we now judge to be most appropriate and least ambiguous, the correlation between MEQ30 scores and change in depression at 4 weeks was 0.38 (p=.066). Although the present study in 24 participants may have been underpowered to detect such an effect at 4 weeks, it is notable that correlations between MEQ30 and GRID-HAMD at the later timepoints were substantially lower (0.17, 0.05, and 0.19 at 3, 6, and 12 months, respectively). This suggests that MEQ30 is not a reliable predictor of enduring antidepressant effects of psilocybin.

The study protocol allowed enrolling individuals who were participating in concurrent psychotherapy if the type and frequency of the therapy had been stable for at least two months prior to screening and was expected to remain stable during participation in the study. Participants were neither discouraged nor encouraged to seek out concurrent psychotherapy during the study. At each long-term follow-up timepoint participants were asked if they had received psychotherapy since receiving psilocybin. At 4 weeks and 3, 6, and 12 months the number of participants (and percentage of the group) who reported receiving psychotherapy since receiving psilocybin was 8 (33.3%), 9 (37.5%), 10 (41.7%), and 10 (41.7%), respectively. There were no significant differences in 12-month GRID-HAMD between participants who did vs. did not report having any psychotherapy at any time during the follow-up period. There were also no differences between these groups in baseline GRID-HAMD, age, sex, number of years with depression, or duration of current depressive episode.

References

Carbonaro TM, Bradstreet MP, Barrett FS, MacLean KA, Jesse R, Johnson MW, Griffiths RR (2016) Survey study of challenging experiences after ingesting psilocybin mushrooms: Acute and enduring positive and negative consequences. *Journal of Psychopharmacology* 30: 1268-1278.

Davis AK, Barrett FS, May DG, Cosimano MP, Sepeda ND, Johnson MW, Finan PH, Griffiths RR (2021) Effects of psilocybin-assisted therapy on major depressive disorder: a randomized clinical trial. *JAMA Psychiatry* 78(5): 481-489.

Griffiths RR, Johnson MW, Carducci MA, Umbricht A, Richards WA, Richards BD, Cosimano MP, Klinedinst MA (2016) Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: A randomized double-blind trial. *Journal of Psychopharmacology* 30(12): 1181-1197.

Griffiths RR, Johnson M, Richards W, Richards B, McCann U, Jesse R (2011) Psilocybin occasioned mystical-type experiences: immediate and persisting dose-related effects. *Psychopharmacology (Berl)* 218(4): 649-665.

Griffiths RR, Richards WA, McCann U, Jesse R (2006) Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. *Psychopharmacology* 187(3): 268-283.

Halpern J H, Lerner A G, Passie T (2016) A review of hallucinogen persisting perception disorder (HPPD) and an exploratory study of subjects claiming symptoms of HPPD. In: Halberstadt AL, Vollenweider FX, and Nichols DE (eds) *Behavioral Neurobiology of Psychedelic Drugs.* Berlin, Heidelberg: Springer, pp. 333-360.

Kroenke K, Spitzer R L (2002) The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric Annals* 32: 509-515.

Posner K, Brent D, Lucas C, Gould M, Stanley B, Brown G, Fisher P, Zelazny J, Burke A, Oquendo M (2008) Columbia-suicide severity rating scale (C-SSRS). New York, NY: Columbia University Medical Center 10.

Posner K, Brown G K, Stanley B, Brent D A, Yershova K V, Oquendo M A, Currier G W, Melvin G A, Greenhill L, Shen S (2011) The Columbia–Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry* 168(12): 1266-1277.

Ross S, Bossis A, Guss J, Agin-Liebes G, Malone T, Cohen B, Mennenga SE, Belser A, Kalliontzi K, Babb J, Su Z, Corby P, Schmidt BL (2016) Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: A randomized controlled trial. *Journal of Psychopharmacology* 30(12): 1165-1180.

Table S1. Depression and suicidal ideation scores at baseline through 12 months.
Post-hoc comparisons between baseline and each followup timepoint are shown, as are the ANOVA results including effect sizes. N=24 unless otherwise noted.

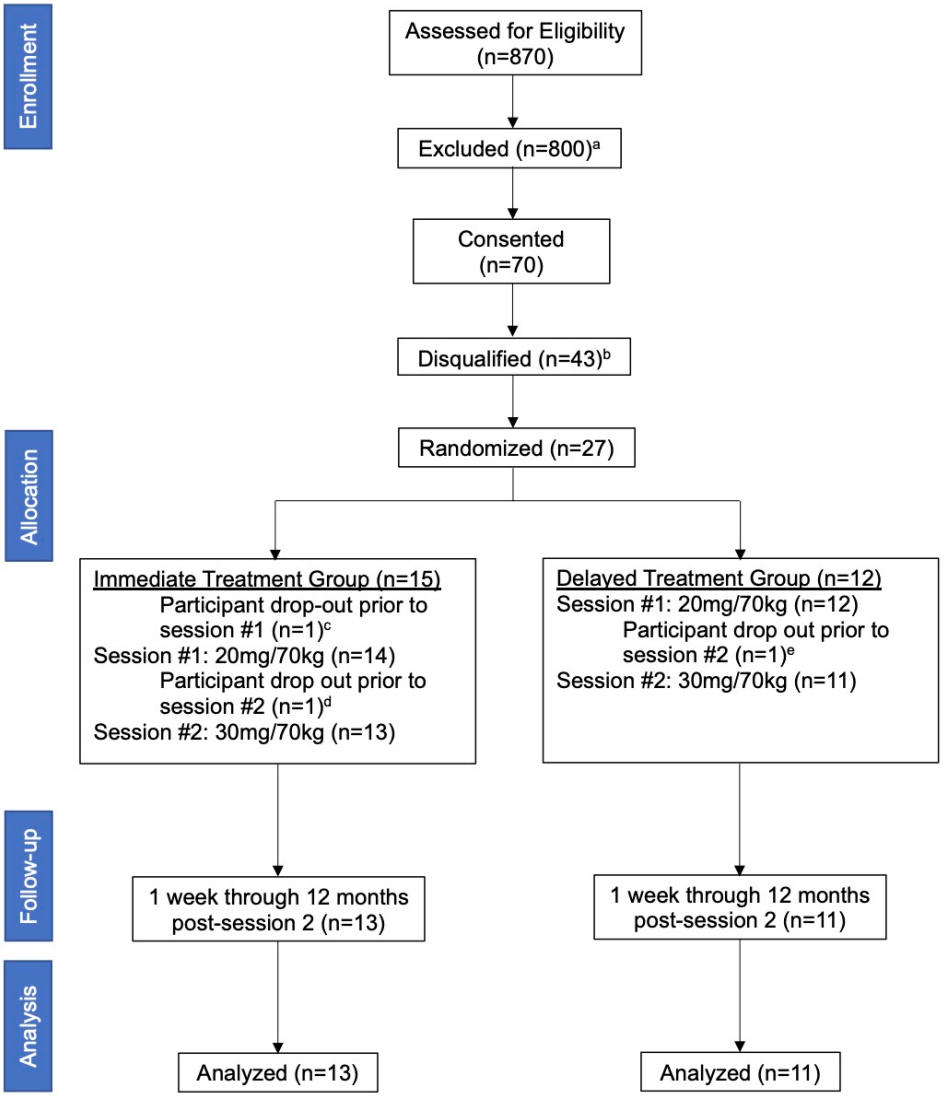
	Baseline	1 week	4 weeks	3 months	6 months	12 months	dF	*F*-stat [a]	Effect size
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)			η_p^2 (90% CI) [b]
Depression									
GRID-HAMD	22.8 (3.9)	8.7 (3.9)†	8.9 (7.4)†	9.3 (8.8)†	7.0 (7.7)†	7.7 (7.9)†	4.1	36.7**	0.61 (0.50-0.67)
QIDS-SR	16.7 (3.5)	5.8 (5.4)†	6.0 (5.7)†	6.2 (5.3)†	5.9 (6.0)†	5.9 (5.8)†	4.3	37.7**	0.62 (0.51-0.68)
BDI-II	33.1 (8.4)	10.1 (11.4)†	9.3 (11.3)†	9.0 (11.3)	8.9 (13.0)†	9.4 (14.2)†	4.0	40.5**	0.64 (0.52-0.70)
Suicidal ideation (SI)									
CSSRS SI item [c]	1.2 (1.2)	.3 (.7)	.2 (.7)	.5 (1.0)	.6 (1.1)	.6 (1.1)	3.5	4.6*	0.17 (0.04-0.27)

[a] *p<.01, **p<.001
[b] Effect sizes (η_p^2, partial eta squared) are shown when *F*-tests are significant; ^small effect (>.01), ^^medium effect (>.06), ^^^large effect (>.14)
[c] N=23 due to missing data
† Indicates that value was significantly different from baseline; values with the same symbol in each row are not statistically different from one another using post-hoc mean pairwise comparisons with Bonferroni correction
M: mean; SD: standard deviation
GRID-HAMD: GRID-Hamilton Depression Rating Scale. Range of scores 0-52; higher scores indicate more severe depression
QIDS-SR: Quick Inventory of Depression Symptoms. Range of scores 0-27; higher scores indicate more severe depression
BDI-II: Beck Depression Inventory II. Range of scores 0-63; higher scores indicate more severe depression
CSSRS: Columbia Suicide Severity Rating Scale. Data are highest suicidal ideation (SI) rating under the "suicidal ideation" assessment (0=no passive or active SI; 1=passive SI; 2=nonspecific active SI; 3=active SI with any methods but no intent or plan; 4=active SI with some intent, without specific plan; 5=active SI with specific plan and intent)

Table S2. Participant ratings of effects attributed to psilocybin session experiences at the 12-month follow-up (N=24).
Data on attitudes, mood, social effects, behavior, and spirituality are means expressed as percentage of maximum possible score, with 1 SD shown in parentheses (N=24).

Persisting Effects Questionnaire Subscales	
Positive attitudes about life	67 (22.1)
Negative attitudes about life	6.7 (10)
Positive attitudes about self	56.8 (21.9)
Negative attitudes about self	5.2 (8.1)
Positive mood changes	59.5 (27.1)
Negative mood changes	4.3 (10.5)
Positive social effects	56.4 (23.9)
Antisocial/negative social effects	1.9 (4.3)
Positive behavior change	75 (27.2)
Negative behavior change	0.8 (4.1)
Increased spiritualty	50.7 (25.9)
Decreased spirituality	1.4 (4.5)

Figure S1. CONSORT diagram of participant flow.



[a] After completing the prescreening questionnaire, people were deemed ineligible if they were currently using antidepressant medication (n=157); lived outside reasonable commuting distance (n=161); did not meet criteria for the magnetic resonance imaging scans (n=99); had a first- or second-degree relative with a diagnosis of schizophrenia spectrum, bipolar I or II, or other psychotic disorder (n=77); had a recent history of substance use disorder (n=50); opted out of in-person screening (n=38); were not in a current depressive episode (n=37); were more than 25% beyond the upper or lower range of recommended body weight (n=32); had a medically significant suicide attempt (n=30); had lifetime hallucinogen use that exceeded the exclusion threshold (n=30); if major depressive disorder (MDD) was not primary psychiatric diagnosis (n=18); if they had a medical exclusion (n=11); had exclusionary use of non-serotonergic psychoactive medication (n=11); or failed to respond to electroconvulsive therapy during current depressive episode (n=4). Forty-five people were ineligible for other reasons.

[b] People were deemed ineligible during in-person screening if they had a psychiatric condition judged to be incompatible with establishment of rapport or safe exposure to psilocybin (n=17); did not have confirmed DSM-5 diagnosis of MDD (n=7); had a recent history of moderate to severe substance use disorder (n=5); were at high risk for suicidality (n=3); disagreed with study procedures (n=3); had a baseline GRID Hamilton Depression Rating Scale score lower than the eligibility threshold of 17 (n=2); had cardiovascular conditions (n=2); had lifetime hallucinogen use that exceeded the exclusion threshold (n=2); were currently taking serotonergic medication (n=1); or were more than 25% beyond the upper and lower range of recommended body weight (n=1).

[c] Dropped out of the study due to anticipatory anxiety about the upcoming first psilocybin session.

[d] Dropped out of study due to sleep difficulties. Sleep difficulties were also reported at screening, and it was not clear whether sleep difficulties were exacerbated by the intervention.

[e] Participant showed a marked reduction in depression symptoms immediately following the first psilocybin session and chose not to proceed with the intervention.

Figure S2. Decrease in the Quick Inventory of Depressive Symptoms (QIDS) scores over time from baseline through the 12-month follow-up (N=24).
Data points are means and brackets are ±1SD; lower brackets are truncated at QIDS scores of 0. QIDS was 16.7 (3.5) at baseline, 5.8 (5.4) at 1 week, 6.0 (5.7) at 4 weeks, 6.2 (5.3) at 3 months, 5.9 (6.0) at 6 months, and 5.9 (5.8) at 12 months post-treatment. All timepoints are significantly different from baseline ($p<0.001$). Cohen *d* effect size is shown for each timepoint. Cohen *d* (95% CI) was 2.4 (1.6-3.1) at 1 week, 2.3 (1.5-3.0) at 4 weeks, 2.3 (1.5-3.1) at 3 months, 2.2 (1.5-2.9) at 6 months, and 2.3 (1.5-3.0) at 12 months.

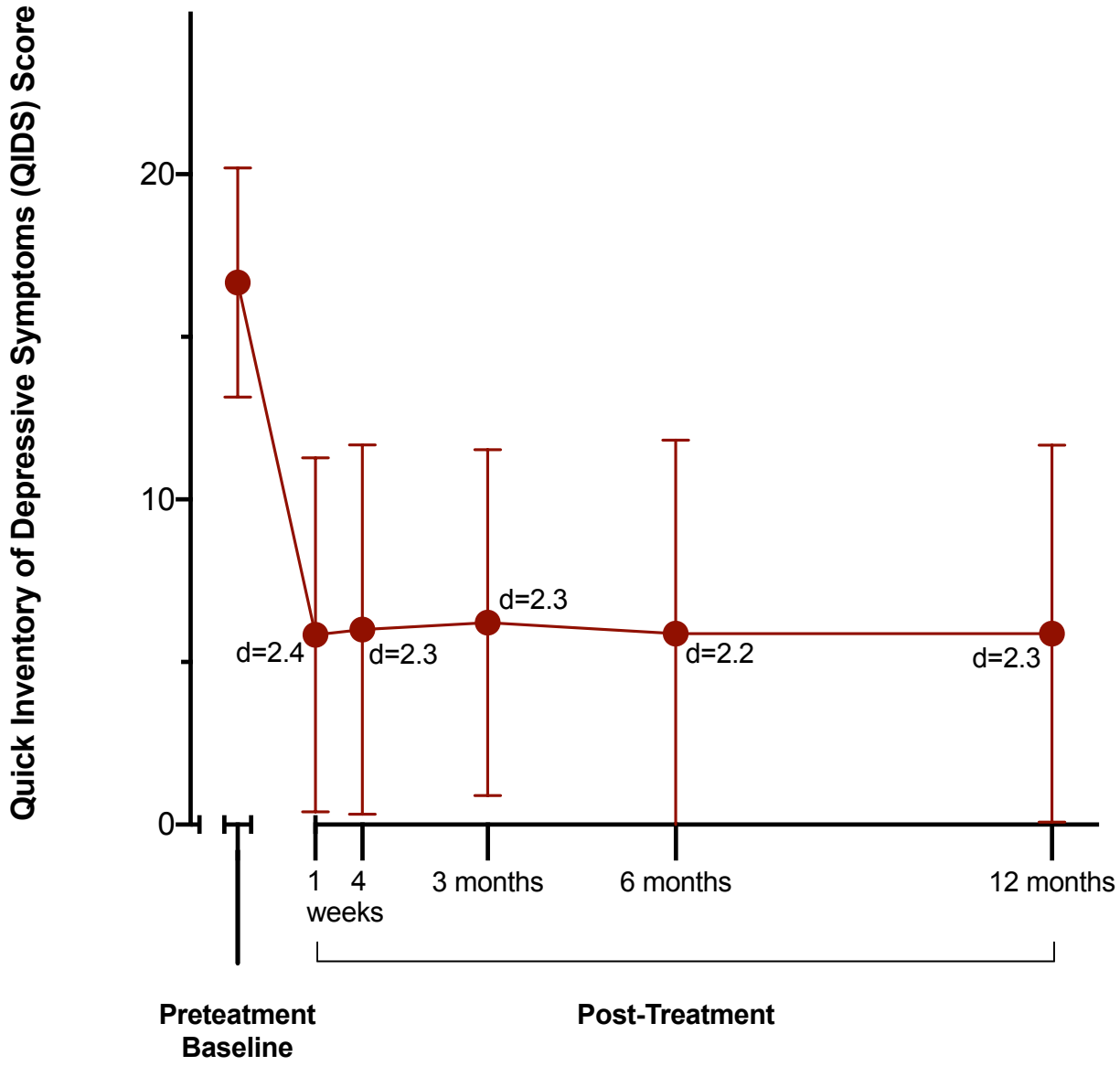


Figure S3. Decrease in the Beck Depression Inventory II (BDI-II) scores over time from baseline through the 12-month follow-up (N=24).
Data points are means and brackets are ±1SD; lower brackets are truncated at BDI-II scores of 0. BDI-II was 33.1 (8.4) at pretreatment baseline, 10.1 (11.4) at 1 week, 9.3 (11.3) at 4 weeks, 9.0 (11.2) at 3 months, 8.9 (13.0) at 6 months, and 9.4 (14.2) at 12 months post-treatment. All timepoints are significantly different from baseline ($p<0.001$). Cohen *d* effect size is shown for each timepoint. Cohen *d* (95% CI) was 2.3 (1.5-3.1) at 1 week, 2.4 (1.6-3.2) at 4 weeks, 2.3 (1.5-3.1) at 3 months, 2.2 (1.4-2.9) at 6 months, and 2.0 (1.3-2.7) at 12 months.