Real-World Psilocybin Therapy for Treatment-Resistant Depression: a Retrospective Observational Study

Johannes Jungwirth

j.mov.com@gmail.com

Psychiatric University Clinic Zurich and University of Zurich, Switzerland https://orcid.org/0009-0008-8142-5874

Samuel Westenhöfer
Helena Aicher
Barbora Provaznikova
Golo Kronenberg
Erich Seifritz
Susanne Prinz
Sebastian Olbrich

University Zürich https://orcid.org/0000-0001-5557-4878

Article

Keywords:

Posted Date: December 10th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-8079137/v1

Additional Declarations: Yes J.J. received honoraria for lectures and advisory board participation from Janssen Switzerland and honoraria for a talk from Schwabe Pharma. H.D.A. has received advisory fees from Mind Medicine Inc., Open Foundation, California Institute of Integral Studies, and honoraria from the Federal Office of Public Health and from the Swiss Medical Society for Psychedelic Therapy. E.S. has received honoraria for educational lectures, consulting fees and unrestricted educational grants from Angelini, Boehringer Ingelheim, Lundbeck, Mepha, OM Pharma, Otsuka, Ricordati, Roche Pharma, Sandoz, Schwabe Pharma, Janssen Cilag, Servier, Sunovion, Takeda and Vifor. S.O. received honoraria for lectures and advisory board participation from Janssen Switzerland and honoraria for talks from

Schwabe Pharma and Indorsia. S.O. is co-founder of DeepPSY AG. All other authors declare no conflict of interest.

Abstract

Psilocybin has demonstrated promising antidepressant effects in depression and treatment-resistant depression (TRD) in controlled clinical trials. However, its effectiveness and safety in real-world therapeutic settings remain largely unknown. Although psilocybin is not yet approved as an antidepressant treatment, Switzerland's unique legal framework allows its limited medical use for TRD. We conducted a retrospective analysis of medical records from 19 TRD patients treated with psilocybin (20–35mg) across one to four dosing sessions at the Psychiatric University Hospital Zurich. Depression severity was assessed using the Montgomery–Åsberg Depression Rating Scale (MADRS) and the Beck Depression Inventory II (BDI). Changes from baseline to interim and post-treatment were analyzed, including response, remission, and the reliable change index. MADRS scores significantly decreased from baseline (M = 30.78) to post-treatment (M = 19.89), with a large effect size (Hedges' g = 1.37, p < .001). BDI scores also decreased significantly (M = 32.33 to M = 23.28), with a large effect (r = .80, p = .003). Response and remission rates were 33.3% and 22.2% (MADRS), and 27.8% and 27.8% (BDI). No additive effect of multiple dosing was found. No serious adverse events occurred. We observed a significant and clinically meaningful reduction in depressive symptoms after psilocybin treatment, with response and remission rates below those reported in previous trials. Although observational and limited by its sample size, this study provides some of the first real-world evidence on psilocybin. Larger, prospective trials are needed to confirm our findings and identify predictors to increase treatment effectiveness.

Introduction

Depressive disorders affect an estimated 330 million people globally and rank among the leading contributors to disease burden, with major depressive disorder (MDD) as the most prevalent subtype (1, 2). Approximately half of the individuals with MDD who complete at least two antidepressant trials do not achieve remission (3). This group is commonly defined as suffering from treatment-resistant depression (TRD) (4). Apart from treatment resistance, patients with TRD show a significantly lower quality of life and higher all-cause mortality, underscoring the urgent need for novel therapeutic options (5, 6).

In this context, psilocybin is emerging as a promising candidate. It primarily acts as a 5-HT2A receptor agonist, inducing dose-dependent alterations in cognition, perception, and affect (7). While the specific mechanisms underlying the antidepressant effect of psilocybin are not fully understood, they are likely involving an interaction between neural plasticity, connectivity changes, and increased psychological flexibility (8). Over the past decade, several randomized controlled trials (RCTs) have demonstrated the efficacy of psilocybin in the treatment of MDD, including TRD (9). Current data suggests that psilocybin is generally safe and well tolerated when administered with psychological support in a supervised clinical setting (10).

Despite these encouraging findings, RCTs on psilocybin have faced methodological challenges. First, they are conducted under strict eligibility criteria. Patients suffering from severe treatment resistance, comorbidities, suicidality or specific medication histories are thus often excluded. Consequently, the generalizability of results from these tightly controlled studies to the broader clinical populations remains uncertain (11). Second, the pronounced subjective effects of psilocybin make effective blinding difficult in both patients and raters, increasing the risk of strong confirmation and expectancy biases in treatment groups, and potential frustration in control groups. This may inflate the estimation of psilocybin`s therapeutic efficacy (12). Taken together, these limitations suggest that, despite their clear value, RCTs alone are insufficient to capture psilocybin's impact on clinical practice.

Another unresolved question is whether current psilocybin treatment paradigms are clinically optimal with respect to both efficacy and durability. Most clinical trials rely on a single medium-to-high dose of psilocybin (13–15). However, a recent follow-up study by Goodwin et al. (16) indicated that relapses may begin to occur as early as three months after participants received one dose of 25 mg psilocybin, challenging the durability of the antidepressant effect (16). A solution to prevent these relapses may lie in multiple dosing sessions. Yet, data on multiple psilocybin sessions are conflicting and scarce. A meta-analysis by Yu et al. (17) suggested a stronger antidepressant effect of two psilocybin dosing sessions compared to one, and in a small scale RCT Rosenblat et al. (18) observed that up to three flexibly planned dosing sessions were associated with further reductions in depressive symptoms in patients with TRD. In contrast, a meta-analysis by Salvetti et al. (19) found no significant difference between one and two psilocybin dosing sessions. Collectively, the uncertainty of response durability and the emerging evidence supporting multidose paradigms, along with the methodological limitations of current RCTs, emphasize the necessity to examine psilocybin treatment under flexible clinical conditions. Switzerland offers a unique legal framework to explore psilocybin treatment for TRD (20). From 2021, psilocybin can be prescribed under a so-called "limited medical use exemption" for several treatment-resistant indications issued by the Federal Office of Public Health (21). In contrast to typical trial conditions, such treatment can be flexibly scheduled. Patients can undergo multiple psilocybin sessions, and antidepressant medication can be continued, providing ideal conditions to explore the feasibility of psilocybin treatment. To our knowledge, there have only been two studies exploring the effects of psilocybin in a comparable setting. The first used a sample of eight patients to explore the long-term effects of psilocybin on alleviating distress associated with life-threatening diseases in Canada (22). The second, by Calder et al. (21), primarily examined the relationship between the acute subjective drug effects and the change in depression scores with both lysergic acid diethylamide (LSD) and psilocybin. While they included patients who were given up to five dosing sessions, they did not explore the course of depression across these dosing sessions.

This present investigation therefore aims to examine the clinical course of TRD patients, receiving up to four psilocybin dosing sessions in a real-world clinical setting. We retrospectively assess changes in depressive symptoms as well as tolerability of repeated psilocybin administrations. We hypothesize that psilocybin treatment will significantly reduce depressive symptoms and that each additional dosing session will further reduce depressive symptoms. By evaluating the efficacy and safety of multiple

psilocybin dosing sessions in a real-world clinical setting, we hope to inform future trials as well as the formation of clinical treatment plans as regulatory approval processes are expected in the upcoming years.

Methods

Study Design

This retrospective observational study was conducted at the University Hospital of Psychiatry Zurich, a multi-site psychiatric institution in Switzerland. Data were extracted from electronic health records in 2025. Ethical approval was granted by the Cantonal Ethics Committee Zurich (BASEC-Nr. 2025 – 00559). Prior to treatment, all patients provided written informed consent for the use of anonymized clinical data in research.

Participants

We included 19 patients who underwent psilocybin treatment at the Second Opinion Consultation Service for Depression, a specialized outpatient consultation and treatment service for TRD patients. Patients were referred to the consultation service by external and internal psychiatrists, psychotherapists, or general practitioners. Participants were between 27 and 67 years old and underwent psilocybin treatment between 02.2023 and 05.2025. All had been diagnosed with a treatment-resistant mild to severe depressive episode by trained psychiatrists according to ICD-10 criteria. To be categorized as treatment-resistant, patients had to have undergone at least two unsuccessful antidepressant trials. Patients were required to be in stable physical health and not showing any contraindications as determined by medical history, physical examination, routine laboratory tests, cranial magnetic resonance imaging (cMRI), electrocardiogram (ECG), and electroencephalogram (EEG).

Exclusion criteria were a personal or first-degree family history of psychotic or bipolar disorders. Patients were also excluded if they presented with uncorrected hypertension, a history of stroke, intracerebral hemorrhage, or aneurysm (intracranial, thoracic, abdominal, or peripheral), recent cardiovascular events (within eight weeks), increased intracranial pressure, or inadequately treated hyperthyroidism. Individuals who were pregnant, breastfeeding, or planning pregnancy during the treatment period were not eligible. A summary of patient demographics and clinical characteristics is given in Table 1.

Table 1
Participant demographics and clinical characteristics

Variable	Yes *n* (%)	Details
Demographics		
Age (years)		*N* = 19; *M* = 47.5, *SD* = 12.7, Min = 27.0, Max = 67.0
Female	6 (31.6%)	
Ethnicity		European 18 (94.7%)
		Non-European 1 (5.3%)
Prior psychedelic experience	5 (26.3%)	
Duration of current depressive episode (at baseline in years)		≥ 15 (4 (21.1%))
		8–14 (3 (15.8%))
		3–7 (9 (47.4%))
		≤ 2 (3 (15.8%))
Clinical Characteristics		
MADRS baseline		*N* = 19; *M* = 30.8, *SD* = 9.0, Min = 15.0, Max = 43.0
BDI baseline		*N* = 19; *M* = 32.3, *SD* = 12.3, Min = 14.0, Max = 52.0
Comorbidities (at baseline)		Suspected Personality Accentuation (6 (31.6%))
		Anxiety Disorders (5 (26.3%))
		Post-traumatic stress disorder (3 (15.8%))
		Attention-deficit hyperactivity disorder (2 (10.5%))
		Mental and behavioral disorders due to psychoactive substance use (2 (10.5%))
		Obsessive-compulsive disorder (2 (10.5%))
		Disorders of adult personality and behavior (1 (5.3%))
		Bulimia nervosa (1 (5.3%))

Notes. Values are *n* (percentage of all patients) unless otherwise specified. Continuous variables report *N*, *M*, *SD*, minimum (Min), and maximum (Max). Abbreviations: MADRS = Montgomery-Åsberg Depression Rating Scale; BDI = Beck Depression Inventory - II.

Variable	Yes *n* (%)	Details
		Pain disorder (1 (5.3%))
		Aspergers Syndrome (1 (5.3%))
Treatment History		
Antidepressant trials (at baseline)		*N* = 19; *M* = 5.0, *SD* = 2.8, Min = 2.0, Max = 14.0
Augmentation (at baseline)	6 (31.6%)	Atypical antipsychotic (5 (26.3%))
		Lithium (2 (10.5%))
Augmentation attempts (at baseline)		> 2 (7 (36.8%))
		2 (3 (15.8%))
		1 (6 (31.6%))
		0 (3 (15.8%))
Prior esketamine treatment	7 (36.8%)	
Antidepressant medication (at baseline)	17 (89.5%)	SARI (7 (36.8%))
		SSNRI (7 (36.8%))
		SSRI (6 (31.6%))
		MAO-Inhibitors, TCAs and others (3 (15.8%))
		SMS (2 (10.5%))
Other psychiatric medication (at baseline)		No other medication (12 (63.2%))
		Benzodiazepine (3 (15.8%))
		Anticonvulsant (2 (10.5%))
		Hypnotic (1 (5.3%))
		Typical antipsychotic (1 (5.3%))
Psychotherapy (at baseline)	15 (78.9%)	Cognitive behavioral therapy (11 (57.9%))
		Other (3 (15.8%))
		Psychoanalysis (3 (15.8%))

Notes. Values are *n* (percentage of all patients) unless otherwise specified. Continuous variables report *N*, *M*, *SD*, minimum (Min), and maximum (Max). Abbreviations: MADRS = Montgomery-Åsberg Depression Rating Scale; BDI = Beck Depression Inventory - II.

Variable	Yes *n* (%)	Details
		Uknown/Missing (4 (21.1%))
Psilocybin treatment characteristics		
Psilocybin dose categories (instances N = 40)		20 mg (1 (2.5%))
		25 mg (27 (67.5%))
		30 mg (10 (25%))
		35 mg (2 (5%))
Number of sessions (instances N = 40)		1 Session (19 (100%))
		2 Sessions (13 (68.4%))
		3 Sessions (5 (26.3%))
		4 Sessions (3 (15.8%))
Notes. Values are *n* (percentage of all patients) unless otherwise specified. Continuous variables report *N*, *M*, *SD*, minimum (Min), and maximum (Max). Abbreviations: MADRS = Montgomery-Åsberg Depression Rating Scale; BDI = Beck Depression Inventory - II.		

Procedure

Treatment Course and Dosing

The treatment procedure assigned to each patient was based on availability as well as clinical judgement. Patients were either treated in a single in- or outpatient setting or in a group inpatient setting and received between one and four dosing sessions. The number of dosing sessions in the single setting was determined by clinical assessment and was individually and dynamically planned and could be a few months apart. In the inpatient group setting, the number of dosing sessions was preplanned to two dosing sessions two weeks apart during a four-week in-patient treatment course. Patients received an individualized dose of orally administered synthetic psilocybin, ranging from 20 to 35 mg. Psilocybin was obtained from Dr. Hysek, Biel, Switzerland and formulated according to Good-Manufacturing-Practice in capsules containing 5 mg.

Preparation and Medication Management

Prior to preparation, patients underwent psychiatric assessment including biographical history, psychometric testing, and medical examinations. Patients then completed at least two initial sessions focused on establishing therapeutic rapport and managing expectations. Antidepressant medication

was generally continued and only paused on the dosing day. Medication with 5-HT2 receptor blocking properties were paused at least three half-lives prior to the dosing session. More information on medication management can be found in the supplement.

Dosing Session

Patients received psilocybin in a calm and specifically arranged room at the University Hospital of Psychiatry Zurich. In the single setting, at least one specifically trained therapist was present during the whole duration of the therapy session. In the group setting, at least two therapists were present during the whole duration of the session. At the end of dosing day, patients were given a questionnaire to assess adverse events. Patients were also encouraged to write a report about the experience.

After Dosing Session

Following dosing day, patients participated in a series of integration meetings aimed at reflecting the psychedelic experience, discussing emerging insights, and supporting transfer to daily life.

Detailed information on preparation, dosing day, integration, and patient flow can be found in the supplement and in supplementary Table S1.

Measures

Depression Outcomes

The primary outcomes were scores on the Beck Depression Inventory II (BDI) (24) and the Montgomery–Åsberg Depression Rating Scale (MADRS) after the final dosing session (25). Assessments were made according to clinical relevance before treatment, between dosing sessions and after the last dosing session.

Secondary outcomes included categorical indicators of clinical improvement. Specifically, response was defined as a ≥ 50% reduction from baseline; remission was defined as a total score reduction to ≤ 13 on the BDI and to ≤ 12 on the MADRS; and reliable change defined by the Reliable Change Index (RCI).

Adverse Events (AEs)

AEs were a secondary outcome collected through a questionnaire containing 16 commonly occurring AEs during psilocybin treatment (Table 2). Patients were asked to indicate whether they had experienced any of these AEs or not. Questionnaires were given to patients directly after the dosing session and were worded to capture AEs during and after treatment. Serious AEs were monitored as part of routine clinical procedures and documented in the medical record.

AEs were a secondary outcome collected through a questionnaire containing 16 commonly occurring AEs during psilocybin treatment (Table 2). Patients were asked to indicate whether they had experienced any of these AEs or not. Questionnaires were given to patients directly after the dosing session and were worded to capture AEs during and after treatment. Serious AEs were monitored as part of routine clinical procedures and documented in the medical record.

Statistical Analysis

The primary objective of this study was to quantify the improvement of depression scores in patients who underwent psilocybin treatment. On all analyses, baseline was determined as the closest measurement prior to the first dosing session, with both BDI and MADRS assessed on the same day. If assessments were not done on the same day, the next assessment was chosen.

Primary Endpoint Analyses

For the paired t-test, Wilcoxon signed-rank test, multiple regression, remission, response, and RCI the first follow-up measure within 42 days after the final dosing session was used as post-treatment measurement.

To assess the changes in depressive symptoms from baseline to post-treatment, a paired t-test was conducted. To determine the effect size of the reduction, Hedges' g was calculated (26). Assessments of normality (Q–Q plot and Shapiro–Wilk) indicated some deviation in the difference scores on the BDI. A Wilcoxon signed-rank test was thus conducted, and a rank-biserial correlation was calculated to determine the size of the reduction. For transparency purposes, and since the results were nearly identical, the corresponding t-test and effect size are reported in the Supplementary Materials.

To account for a possible bias in the assessment timing, since post-treatment scores were collected at different time-points, a multiple linear regression model was fitted to predict post-treatment scores using baseline scores and the number of days since treatment as predictors. Results can be found in the supplement

To assess whether group-level changes were clinically meaningful, remission (MADRS ≤ 12; BDI ≤ 13), response (≥ 50% reduction), and the RCI were calculated (27, 28). Thresholds for clinically reliable improvement and clinically reliable deterioration were set to 1.96 and − 1.96 respectively, corresponding to a 95% confidence interval (CI).

Changes in Depressive Symptoms Across Sessions

To assess the effects across multiple psilocybin dosing sessions, a linear mixed-effects model was fitted using the lm4 package (29). The outcome variable for the linear mixed model was derived by selecting the first available score recorded within 42 days after each dosing session. The model included sessions as a factor from 0 (baseline) to 4 (fourth dosing session) as a fixed effect and participants as random intercepts to account for baseline differences among individuals. Estimated marginal means

(EMMs) were calculated and pairwise compared to inspect differences between sessions. Pairwise comparison was adjusted using the Tukey method.

Adverse Events

AEs were summarized by type, first, based on how many participants exhibited them at least once, and second, on how many times it occurred in all dosing sessions.

All analyses were conducted using R (version 4.5.0) and RStudio (version 2025.05.1 + 513). Statistical tests used $p < 0.05$, two-tailed, to determine statistical significance. A full list of the key packages, assumptions tested and further explanations of RCI can be found in the supplement.

Results

Primary Endpoint Analyses

The final sample consisted of 19 patients. All 19 had a baseline assessment and completed the first dosing session. Thirteen completed two dosing sessions, five completed three dosing sessions, and three completed four dosing sessions. For the primary endpoint analyses one participant did not have a valid endpoint assessment on the MADRS and another did not have a valid assessment on the BDI, leading to a final $n = 18$ on both instruments. A full participant flow is provided in the Supplement.

The MADRS scores significantly decreased from baseline ($M = 30.78$; $SD = 9.02$) to after the last dosing session ($M = 19.89$; $SD = 9.65$), $t(17) = 6.067$, $p < .001$; Fig. 1A). The mean reduction was 10.89 points, 95% CI [7.10, 14.68], with a very large effect size (*Hedges' g* = 1.37). A Wilcoxon signed-rank test was conducted to compare BDI scores before ($M = 32.33$, $SD = 12.25$) and after the last dosing session ($M = 23.28$, $SD = 13.88$). A significant reduction in depressive symptoms was found ($V = 153.5$, $p = .003$; Fig. 1B). The mean reduction in BDI scores was 9.06 points, 95% CI [3.48, 14.63]. Again, the effect size was very large ($r = .80$). These results indicate that patients improved across the treatment period across both clinician- and self-rated measures.

Out of 18 patients, 6 (33.3%) and 5 (27.8%) were classified as responders on the MADRS and BDI respectively. The mean percentage reduction was 35.6% on the MADRS and 26.8% on the BDI. In total, 4 patients (22.2%) on the MADRS and 5 patients (27.8%) on the BDI met remission criteria at the end of treatment period. On the MADRS, 10 (55.6%) patients showed reliable improvement, 8 patients (44.4%) showed no reliable change, and no patient showed reliable deterioration. The threshold for a reliable change was 8.21 points. On the BDI, 8 patients (44.44%) showed reliable improvement, 10 patients (55.6%) showed no reliable change, and none showed reliable deterioration. The threshold for a reliable change was 9.5 points.

Changes in Depressive Symptoms Across Sessions

To assess changes across sessions, a linear mixed model was fitted. After removing all invalid assessments, in total 50 were left on the MADRS and 51 were left on the BDI. At baseline, the mean MADRS score was 31.3 (*SD* = 9.1) and the mean BDI score was 33 (*SD* = 12.3). After session one, the mean scores were 17.9 (*SD* = 7.31) on the MADRS and 18.3 (*SD* = 12.6) on the BDI; after session two, 22.6 (*SD* = 11.98) and 24.8 (*SD* = 14.4); after session three, 21 (*SD* = 8.3) and 35.8 (*SD* = 13.2). On the MADRS, two patients were assessed after session four (*M* = 22.5, *SD* = 10.61), and on the BDI, three patients (*M* = 40, *SD* = 8.7). Figures 2A and 2C show the trajectory of depression scores across sessions.

The linear mixed model indicated that MADRS scores were significantly reduced compared to baseline after session one ($b = -13.06$, $SE = 2.39$, $t(28.8) = -5.46$, $p < .001$), session two ($b = -9.22$, $SE = 2.46$, $t(29.1) = -3.74$, $p = .001$), session three ($b = -10.86$, $SE = 3.87$, $t(30.6) = -2.80$, $p = .008$) and session four ($b = -15.17$, $SE = 5.32$, $t(53) = -2.85$, $p = .008$). The intraclass correlations (ICC) showed that 54% of the variance was between patients, indicating that random intercepts were warranted. In total, fixed effects of the model explained 27.6% of the variance in depression scores and the full model (fixed + random effects) explained 66.5% of the variance. On the BDI, significant reductions compared to baseline were found after session one ($b = -12.42$, $SE = 2.71$, $t(29) = -4.57$, $p < .001$) and session two ($b = -9.13$, $SE = 1.05$, $t(53) = -2.80$, $p = .003$). The ICC showed that 68% of the variance was between patients, indicating that random intercepts were warranted. In total fixed effects of the model explained 14.2% of the variance in depression scores and combined with random effects, the full model explained 72.6% of the variance. The complete linear mixed model results for both instruments can be found in Supplementary Tables S2 A and S2 B.

To test whether depression scores differed between sessions one through four, EMMs were calculated and pairwise compared. After correcting for multiple testing, MADRS scores showed a significant reduction between baseline (M = 31.3, 95% CI [26.90, 35.7]) and session 1 (*M* = 18.3, 95% CI [13.26, 23.3]), with an estimated mean difference of 13.06 ($SE = 2.40$, $t(28.7) = 5.44$, $p < .001$; Fig. 2B). A significant reduction was also observed between baseline and session 2 (*M* = 22.1, 95% CI [16.95, 27.2]), with an estimated mean difference of 9.22 ($SE = 2.48$, $t(29) = 3.72$, $p = .007$; Fig. 2B). On the BDI, scores significantly decreased from baseline (*M* = 33, 95% CI [26.90, 39.1]) to session 1 (*M* = 20.6, 95% CI [14.00, 27.2]), with an estimated mean difference of 12.42 ($SE = 2.73$, $t(29.2) = 4.56$, $p < .001$; Fig. 2D). It also significantly decreased after session 2 (*M* = 23.9, 95% CI [17.1, 30.6]) compared to baseline, with an estimated mean difference of 9.13 ($SE = 2.81$, $t(29.2) = 3.25$, $p = .022$; Fig. 2D). Complete tables of all EMMs and comparisons for both instruments are provided in Supplementary Tables S3 A, S3 B, S4 C and S4 B

In summary, the results of the linear mixed model and the pairwise comparison indicate that across instruments there was a significant early decrease in depressive symptoms. This improvement was maintained across later sessions without further significant change.

Adverse Events

In 24 (80%) of the 30 recorded dosing sessions, at least one AE was reported. Patients experiencing an AE typically reported more than one (*M* = 3.53, *SD* = 2.11, Min = 1, Max = 8). The most frequent AEs were fatigue (*n* = 20), headache (*n* = 16), and tearfulness (n = 15). All AEs were transient, and no serious AEs were recorded. Table 4 contains all AEs recorded.

Table 1
Frequency of Adverse Events

Adverse Event	**Patients *n* (%)**	**Sessions *n* (%)**
Fatigue	12 (63.2%)	20 (66.7%)
Headache	9 (47.4%)	16 (53.3%)
Difficulty concentrating	8 (42.1%)	11 (36.7%)
Tearfulness	7 (36.8%)	15 (50.0%)
Anxiety	7 (36.8%)	11 (36.7%)
Nausea	7 (36.8%)	8 (26.7%)
Impaired balance	5 (26.3%)	6 (20.0%)
Dizziness	5 (26.3%)	5 (16.7%)
None	4 (21.1%)	6 (20.0%)
Dry mouth	3 (15.8%)	5 (16.7%)
Palpitations	3 (15.8%)	3 (10.0%)

Note. *N*(patients) = 19; *N*(sessions) = 30; In the patients column every AE is only counted once for each patient.

Discussion

In this retrospective real-world study of patients with TRD, psilocybin was administered between one and four times, flexibly fitted to the course of each patient by clinical decisions. A significant and clinically relevant improvement in depressive symptoms between baseline and after the final dosing session was found on both self-reported and clinician-rated instruments. No serious or sustained AEs occurred. The mean reduction of -10.89 points (MADRS) and of 9.08 points (BDI) observed in this study were slightly lower to the mean reduction of 12.11 points across clinical trials on both TRD and MDD (19). Response and remission numbers of this study were lower than those observed in previous RCTs on MDD, which had up to 70% of patients responding and roughly 50% remitting within weeks of treatment (15, 30, 31). However, they fell within the lower range of recent large scale RCTs with response rates between 37% and 42%, and remission rates between 29% and 25% (13, 14). The effect sizes found in this study were also within the range of prior trials, but smaller in size (19). Taken together, these findings indicate that results from clinical trials, to some extent, translate into routine clinical practice, although outcomes

appear smaller in scale. However, such comparisons warrant caution, as existing studies involved one or two dosing sessions and focused mainly on MDD rather than TRD.

Contrary to our expectation, we did not find an additive effect of multiple psilocybin sessions. On both scales, the strongest improvement in depressive symptoms occurred after the first dosing session, and subsequent dosing sessions did not differ from the initial improvement after the first session. These results could suggest a stabilization or ceiling effect. MADRS scores after all sessions were significantly decreased compared to baseline, but in pairwise comparisons this significance disappeared from session three onwards. On the BDI, only the first two sessions showed significant improvement over baseline, while still showing a trend-level improvement thereafter. Several factors may account for this pattern. First, only a very small number of patients underwent more than two dosing sessions, increasing the uncertainty of the results. Second, we continued treatment even when patients did not respond. In fact, most patients who had more than two dosing sessions did not respond to the treatment. A possible explanation for this lack of response could be that these patients suffered from a more severe form of treatment-resistance. Third, the time between dosing sessions varied greatly between patients, and it may be that there is a specific time window in which an additional dose is most effective. There is currently no data to support this claim, as data on multiple treatment sessions are scarce. Finally, there is the possibility that there is simply no additive or not even a maintaining effect of two or more than two psilocybin sessions. The two meta-analyses conducted on this topic yield inconsistent results; one found that two psilocybin sessions had a greater and longer-lasting effect than one, while the other failed to find a statistically robust difference (17, 19). It should be noted that even though we found no further improvement on a group level, there might still be a merit in multiple psilocybin sessions, as some patients only responded for the first time after the second or third session.

The study at hand is among the first to explore a multidose paradigm as well as the feasibility of psilocybin in a naturalistic clinical real-world setting. The naturalistic setting is especially important for patients with TRD, as they may not be adequately represented in trials, due to strict exclusion criteria (4). By including individuals with complex treatment resistance, multiple comorbidities, numerous unsuccessful antidepressant trials, several augmentation attempts, and chronic depression of many years' duration, our sample reflects a representative TRD population (4). Finally, the multidose paradigm provides unique insights into how psilocybin could be used in practice and may inform the design of future trials, especially regarding the timing and number of dosing sessions.

While this study offers important data on the real-world use of psilocybin, there are limitations that need to be acknowledged. A major limitation was the small sample size, which limits statistical power. This was most apparent in the loss of significance for the effects observed after the third and fourth sessions once corrected for multiple testing. Further, female patients and patients of non-European decent were underrepresented, reducing generalizability. Although data were collected in a longitudinal manner, we only retrospectively reviewed health records, which precludes any causal inference. Adding to this, data was drawn from health records in active clinical service, so some patients may have returned for more dosing sessions after our measurement period. Furthermore, treatment procedures and data collection

were guided by clinical practice rather than a standardized research protocol. As a result, substantial heterogeneity was present across treatment duration, number of dosing sessions, measurement time-points, concomitant psychopharmacotherapy, psychotherapy, and follow-up schedules. This variability paired with the fact that we did not have a control group makes it difficult to determine if changes can be attributed to the psilocybin treatment alone or whether they may be better explained in conjunction with factors such as adjunct medication or the difference in treatment exposure.

While it is important to note that these issues are somewhat common in real-world settings, future research could mitigate some of these limitations by running prospective observational designs comparing psilocybin treatment to other treatment options for TRD such as esketamine or electroconvulsive therapy. Ideally, larger observational studies should control for variables such as concomitant medication and psychotherapy, type of psychotherapy, setting, and treatment duration. Given the resource intensive nature of current psilocybin treatment paradigms (20), future studies should aim to identify biomarkers or other predictors for treatment success that enable patient stratification to increase effectiveness and economic viability.

In sum, this study demonstrates that psilocybin treatment led to a significant and clinically meaningful improvement in depressive symptoms among patients with TRD, replicating findings from clinical trials with a slightly lowered efficacy. We further explored the benefit of multiple psilocybin sessions, finding no additive but likely a sustaining or ceiling effect. This study marks an important first step in investigating the effectiveness of multiple, flexibly scheduled psilocybin sessions in clinical practice. For more definitive conclusions, studies with larger sample sizes are needed.

Declarations

Acknowledgement

The authors thank Eliane Strickler, Godehard Weniger, Yamina Ehrt, Alexandra Georgieva, Guglielmo Lucchese and Tobias Welt for clinical contributions. We thank Juraj Styk for psychotherapeutic supervision. We also thank Matthias Liechti for pharmacological consultation.

Conflicts of interest

J.J. received honoraria for lectures and advisory board participation from Janssen Switzerland and honoraria for a talk from Schwabe Pharma. H.D.A. has received advisory fees from Mind Medicine Inc., Open Foundation, California Institute of Integral Studies, and honoraria from the Federal Office of Public Health and from the Swiss Medical Society for Psychedelic Therapy. E.S. has received honoraria for educational lectures, consulting fees and unrestricted educational grants from Angelini, Boehringer Ingelheim, Lundbeck, Mepha, OM Pharma, Otsuka, Ricordati, Roche Pharma, Sandoz, Schwabe Pharma, Janssen Cilag, Servier, Sunovion, Takeda and Vifor. S.O. received honoraria for lectures and advisory board participation from Janssen Switzerland and honoraria for talks from Schwabe Pharma and Indorsia. S.O. is co-founder of DeepPSY AG. All other authors declare no conflict of interest.

Author Contributions (CRediT taxonomy)

JJ: Conceptualization, Methodology, Study Design (lead), Data Curation, Formal Analysis, Investigation, Therapeutic Intervention, Writing – Original Draft, Visualization. SW: Data Curation, Methodology, Formal Analysis (lead), Visualization, Writing – Original Draft (lead). HDA: Methodology, Investigation, Therapeutic Intervention, Writing – Review & Editing; BP: Investigation, Writing – Review & Editing; GK: Writing – Review & Editing; ES: Writing – Review & Editing; SP: Methodology, Investigation, Therapeutic Intervention, Writing – Review & Editing; SO: Conceptualization, Supervision, Investigation, Therapeutic Intervention, Writing – Review & Editing.

Declaration of use of artificial intelligence (AI):

The AI-based tool ChatGPT (model: GPT-5 Thinking, OpenAI) was used at certain points for support in translation and wording. The content, ideas, structure, argumentation, and selection of the literature originate entirely from the authors. The authors take full responsibility for the text and the work as a whole. All AI-assisted passages were reviewed, revised, and source-verified by the authors. No patient data were entered into any AI system.

Supplementary information is available at MP's website

References

1. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2021 (GBD 2021) Results Depressive Disorders Prevalence 2021 [Internet]. 2024 [cited 2025 Jul 10]. Available from: https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/972c32ffcb1e24b1c40c487c6282ab15
2. Ferrari AJ, Santomauro DF, Aali A, Abate YH, Abbafati C, Abbastabar H, et al. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet. 2024;403(10440):2133–61.
3. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and Longer-Term Outcomes in Depressed Outpatients Requiring One or Several Treatment Steps: A STAR*D Report. American Journal of Psychiatry. 2006;163(11):1905–17.
4. McIntyre RS, Alsuwaidan M, Baune BT, Berk M, Demyttenaere K, Goldberg JF, et al. Treatment-resistant depression: definition, prevalence, detection, management, and investigational interventions. World Psychiatry. 2023;22(3):394–412.
5. Jaffe DH, Rive B, Denee TR. The humanistic and economic burden of treatment-resistant depression in Europe: a cross-sectional study. BMC Psychiatry. 2019;19(1):247.
6. Brenner P, Reutfors J, Nijs M, Andersson TML. Excess deaths in treatment-resistant depression. Ther Adv Psychopharmacol. 2021;11.

7. Vollenweider FX, Preller KH. Psychedelic drugs: neurobiology and potential for treatment of psychiatric disorders. Nat Rev Neurosci. 2020;21(11):611–24.
8. Liebnau J, Betzler F, Kerber A. Catalyst for change: Psilocybin's antidepressant mechanisms—A systematic review. Journal of Psychopharmacology. 2025;39(5):397–415.
9. Metaxa AM, Clarke M. Efficacy of psilocybin for treating symptoms of depression: systematic review and meta-analysis. BMJ. 2024;e078084.
10. Jungwirth J, Bavato F, Quednow BB. Psychedelika und Dissoziativa in der Psychiatrie: Herausforderungen in der Behandlung. Nervenarzt. 2024;95(9):803–10.
11. Hovmand OR, Poulsen ED, Arnfred S, Storebø OJ. Risk of bias in randomized clinical trials on psychedelic medicine: A systematic review. Journal of Psychopharmacology. 2023;37(7):649–59.
12. Hieronymus F, López E, Werin Sjögren H, Lundberg J. Control Group Outcomes in Trials of Psilocybin, SSRIs, or Esketamine for Depression. JAMA Netw Open. 2025;8(7):e2524119.
13. Goodwin GM, Aaronson ST, Alvarez O, Arden PC, Baker A, Bennett JC, et al. Single-Dose Psilocybin for a Treatment-Resistant Episode of Major Depression. New England Journal of Medicine. 2022;387(18):1637–48.
14. Raison CL, Sanacora G, Woolley J, Heinzerling K, Dunlop BW, Brown RT, et al. Single-Dose Psilocybin Treatment for Major Depressive Disorder A Randomized Clinical Trial. JAMA. 2023;330(9):843–53.
15. von Rotz R, Schindowski EM, Jungwirth J, Schuldt A, Rieser NM, Zahoranszky K, et al. Single-dose psilocybin-assisted therapy in major depressive disorder: a placebo-controlled, double-blind, randomised clinical trial. EClinicalMedicine. 2023;56:101809.
16. Goodwin GM, Nowakowska A, Atli M, Dunlop BW, Feifel D, Hellerstein DJ, et al. Results From a Long-Term Observational Follow-Up Study of a Single Dose of Psilocybin for a Treatment-Resistant Episode of Major Depressive Disorder. 2025.
17. Yu CL, Liang CS, Yang FC, Tu YK, Hsu CW, Carvalho AF, et al. Trajectory of Antidepressant Effects after Single- or Two-Dose Administration of Psilocybin: A Systematic Review and Multivariate Meta-Analysis. J Clin Med. 2022;11(4):938.
18. Rosenblat JD, Meshkat S, Doyle Z, Kaczmarek E, Brudner RM, Kratiuk K, et al. Psilocybin-assisted psychotherapy for treatment resistant depression: A randomized clinical trial evaluating repeated doses of psilocybin. Med. 2024;5(3):190–200.e5.
19. Salvetti G, Saccenti D, Moro AS, Lamanna J, Ferro M. Comparison between Single-Dose and Two-Dose Psilocybin Administration in the Treatment of Major Depression: A Systematic Review and Meta-Analysis of Current Clinical Trials. Brain Sci. 2024;14(8):829.
20. Liechti ME, Gasser P, Aicher HD, Mueller F, Hawrot T, Schmid Y. Implementing psychedelic-assisted therapy: History and characteristics of the Swiss limited medical use program. Neuroscience Applied. 2025;4:105525.
21. Aicher HD, Schmid Y, Gasser P. Psychedelika-assistierte Psychotherapie. Die Psychotherapie. 2024;69(2):98–106.

22. de la Salle S, Kettner H, Thibault Lévesque J, Garel N, Dames S, Patchett-Marble R, et al. Longitudinal experiences of Canadians receiving compassionate access to psilocybin-assisted psychotherapy. Sci Rep. 2024;14(1):16524.
23. Calder AE, Rausch B, Liechti ME, Holze F, Hasler G. Naturalistic psychedelic therapy: The role of relaxation and subjective drug effects in antidepressant response. Journal of Psychopharmacology. 2024;38(10):873–86.
24. Kühner C, Bürger C, Keller F, Hautzinger M. Reliabilität und validität des revidierten Beck-Depressionsinventars (BDI-II). Befunde aus deutschsprachigen stichproben. Nervenarzt. 2007;78(6):651–6.
25. Schmidtke A, Fleckenstein P, Moises W, Beckmann H. [Studies of the reliability and validity of the German version of the Montgomery-Asberg Depression Rating Scale (MADRS)]. Schweiz Arch Neurol Psychiatr (1985). 1988;139(2):51–65.
26. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Front Psychol. 2013;4(NOV).
27. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59(1):12–9.
28. Blampied NM. Reliable change and the reliable change index: still useful after all these years? Cogn Behav Therap. 2022;15(3).
29. Bates D, Maechler M, Bolker B, Walker S. lme4: Linear Mixed-Effects Models using “Eigen” and S4. CRAN: Contributed Packages. 2003.
30. Davis AK, Barrett FS, May DG, Cosimano MP, Sepeda ND, Johnson MW, et al. Effects of Psilocybin-Assisted Therapy on Major Depressive Disorder. JAMA Psychiatry. 2021;78(5):481.
31. Carhart-Harris R, Giribaldi B, Watts R, Baker-Jones M, Murphy-Beiner A, Murphy R, et al. Trial of Psilocybin versus Escitalopram for Depression. New England Journal of Medicine. 2021;384(15):1402–11.

Figures

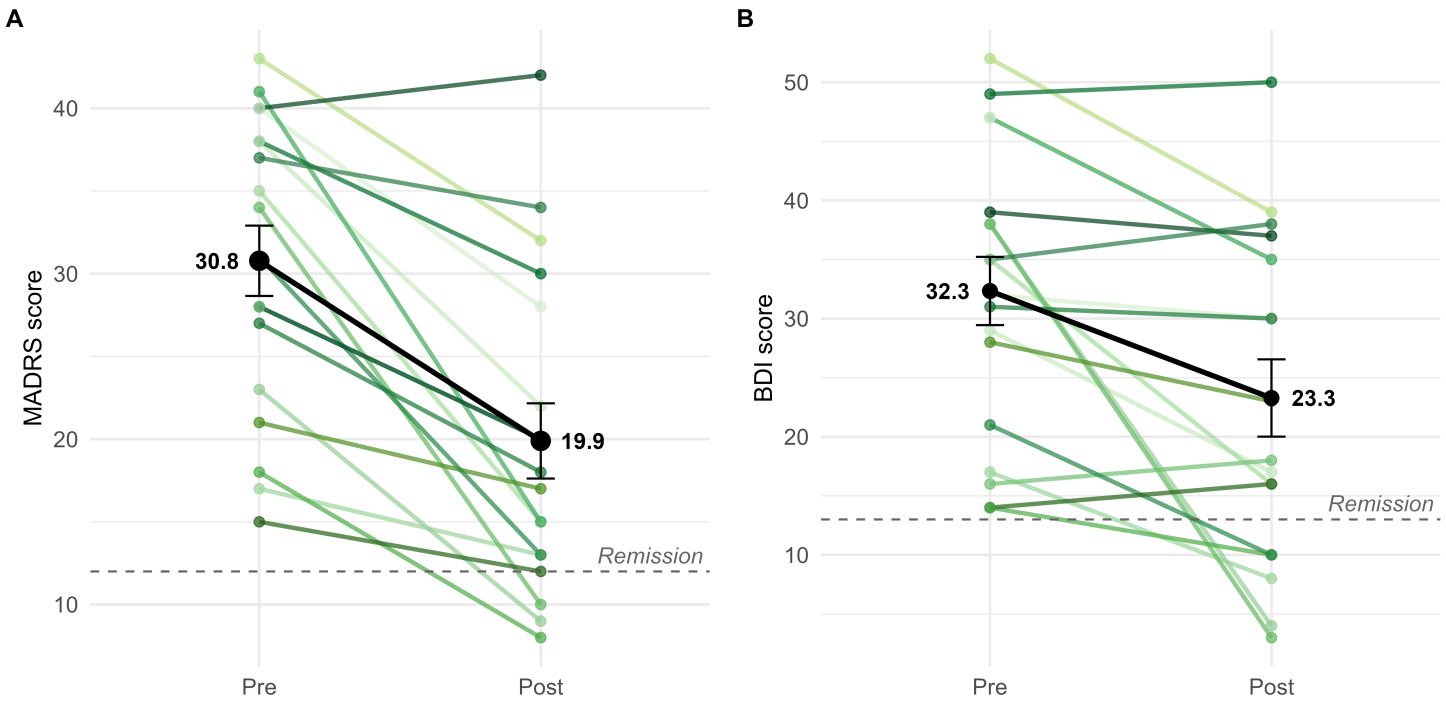

Figure 1

A. Change in MADRS scores from pre- to post-treatment.
The black line and error bars indicate the group mean ± SE. The dashed line marks the remission threshold (MADRS ≤ 12).

B. Change in BDI scores from pre- to post-treatment.
The black line and error bars indicate the group mean ± SE. The dashed line marks the remission threshold (BDI ≤ 13).

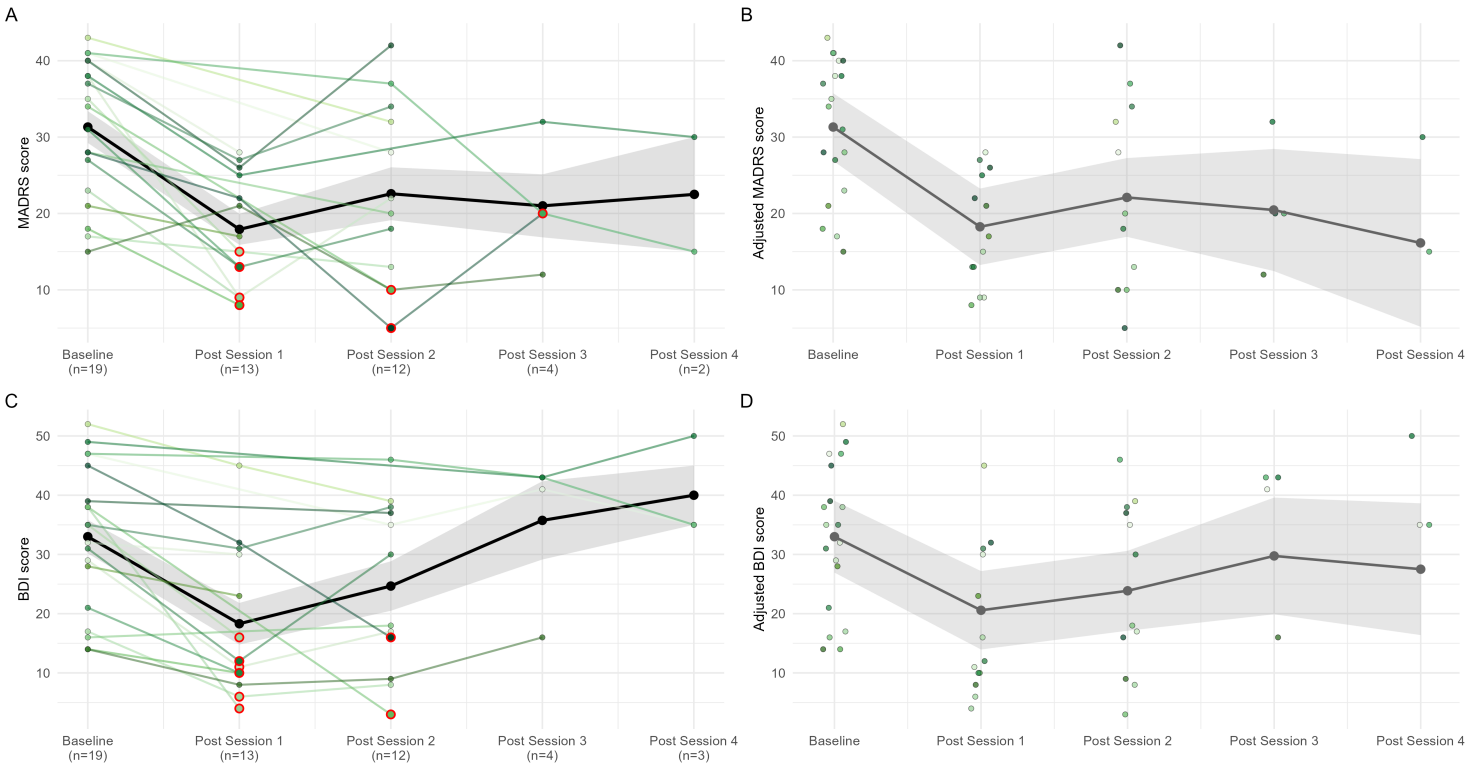

Figure 2

A. Trajectory of clinician-rated depression severity (MADRS) across sessions.

The black line and translucent band indicate the group mean ± SE. Red-outlined dots mark the sessions at which a ≥ 50% reduction from baseline was reached for the first time.

B. Estimated marginal means of clinician-rated depression severity (MADRS) across sessions.

The line and points show model-adjusted means; the shaded band shows the 95% CI. Green colored dots represent participant scores.

C. Trajectory of self-rated depression severity (BDI) across sessions.

The black line and translucent band indicate the group mean ± SE. Red-outlined dots mark the sessions at which a ≥50% reduction from baseline was reached for the first time.

D. Estimated marginal means of self-rated depression severity (BDI) across sessions.

The line and points show model-adjusted means; the shaded band shows the 95% CI. Green colored dots represent participant scores.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx