# Psilocybin for depression: Considerations for clinical trial design

KELLEY C. O'DONNELL[1,2]*, SARAH E. MENNENGA[1,2] and MICHAEL P. BOGENSCHUTZ[1,2]

[1]Department of Psychiatry, New York University School of Medicine, New York, NY, USA
[2]Department of Psychiatry, Bellevue Hospital Center, New York, NY, USA

*Background and aims:* Given the enormous global burden of depressive illness, there is an urgent need to develop novel and more effective treatments for major depressive disorder (MDD). Recent findings have suggested that psychedelic drugs may have a role in the treatment of depressive symptoms, and a number of groups are in the process of developing protocols to study this question systematically. Given the subjective quality of both the psychedelic experience and depressive symptomatology, great care must be taken when designing a protocol to study the clinical efficacy of psychedelic drugs. This study will discuss many factors to consider when designing a clinical trial of psilocybin for MDD. *Methods:* We provide a thorough review of pertinent research into antidepressant clinical trial methodology and review practical considerations that are relevant to the study of psychedelic-assisted treatment for depression. *Results:* We discuss participant selection (including diagnostic accuracy, exclusion criteria, characteristics of the depressive episode, and the use of concurrent medications), study interventions (including dosing regimens, placebo selection, non-pharmacological components of treatment, and the importance of blinding), trial duration, outcome measures, and safety considerations. *Conclusions:* Careful and transparent study design and data analysis will maximize the likelihood of generating meaningful, reproducible results, and identifying a treatment-specific effect. Meeting the highest standards for contemporary trial design may also broaden the acceptance of psychedelic research in the scientific community at large.

*Keywords:* psilocybin, psychedelic, depression, major depressive disorder

## INTRODUCTION

Although clinical research on classic psychedelics [e.g., psilocybin, lysergic acid diethylamide (LSD), and ayahuasca] dates back almost 70 years, its current iteration is relatively new. The number of groups studying the therapeutic value of these drugs is growing, but the clinical data are still quite limited and of highly variable quality. Designing future trials to the highest standards will maximize the likelihood of finally achieving satisfactory answers to basic questions of efficacy and safety.

Given the enormous global burden of depressive illness, the need for new and better interventions is particularly acute (Ferrari et al., 2013). Over the past several years, a number of groups have become interested in studying the utility of classic psychedelics for depression. With respect to psilocybin, this interest began with pilot studies (Carhart-Harris et al., 2016; Grob et al., 2011) and was bolstered by two randomized controlled trials (RCTs) of psilocybin for depression and anxiety in patients with cancer (Griffiths et al., 2016; Ross et al., 2016), and an open-label pilot study in patients with treatment-resistant depression (TRD; Carhart-Harris et al., 2016). Given the positive results of these early studies, larger trials are currently in development, some of which will focus on patients with major depressive disorder (MDD) (ClinicalTrials.gov Identifiers: NCT03181529, NCT03554174, NCT03429075, and NCT03380442).

Although there is a burgeoning scientific and public interest in the mechanisms and therapeutic utility of psychedelic drugs, methodological rigor in studies of MDD is an ongoing challenge (Schatzberg, 2019). In this review, we will discuss a number of factors that are important in psychedelic trial design for MDD. We address factors important for any rigorous study of MDD, as well as factors particularly relevant to clinical trials of psychedelics for MDD. Given the weight of the evidence and our own expertise, we will focus primarily on the psychedelic drug psilocybin.

## SELECTION OF PARTICIPANTS

### Diagnostic accuracy

MDD is a heterogeneous illness, with a high degree of variability in all aspects of the disorder (e.g., age of onset, symptom clusters, symptom severity, episodic vs.

* Corresponding author: Kelley C. O'Donnell, MD, PhD; Department of Psychiatry, New York University School of Medicine, One Park Avenue, 8th Floor, New York, NY 10016, USA; Bellevue Hospital Center, 462 First Avenue, A Building, Room A834, New York, NY 10016, USA; Phone: +1 646 501 2249; Fax: +1 765 227 2054; E-mail: odonnk02@nyulangone.org

unremitting disease course, etc.; Lamers et al., 2016; Power et al., 2016; Sedlackova et al., 2015). As the scientific understanding of the relationship between pathophysiology and phenomenology grows, our nosology is likely to change dramatically. In the meantime, however, the Diagnostic and Statistical Manual (DSM) will continue to guide the selection of research subjects for a drug efficacy trial. Study participants in a depression trial should be patients with a DSM-5 (American Psychiatric Association, 2013) diagnosis of MDD, who are currently in a major depressive episode (MDE).

### Structured interviews

The diagnosis of MDD should be carried out through a structured clinical interview. In the DSM-5 field trials, the interrater reliability for MDD was very low (Cohen's κ = 0.28; by comparison, the κ for PTSD was 0.67, and for alcohol use disorder was 0.40; Freedman et al., 2013). As such, we strongly recommend that clinical researchers make use of standardized diagnostic instruments when screening participants. For many years, the gold standard for this was the Structured Clinical Interview for DSM-IV (American Psychiatric Association, 1994), which had an interrater reliability κ value of 0.66 (Lobbestael, Leurgans, & Arntz, 2011) to 0.80 (Zanarini et al., 2000) for MDD. The interrater reliability of the MDD diagnosis has not yet been published for the SCID-5, but we would expect it to be similar to the SCID-IV, as the diagnostic criteria for MDD were minimally changed in the transition.

When making the diagnosis of MDD, particular attention should be given to clinical features, whether current or historical, that might raise suspicion for a bipolar diathesis. In addition to the potential for a reduced antidepressant response (Sidor & Macqueen, 2011), patients with bipolar disorder who take psilocybin may have an increased risk of psychosis, so a personal or family history of bipolar disorder has been an exclusionary criterion in all psilocybin trials of the modern era, regardless of indication. Patients who are diagnosed with bipolar disorder frequently present with an MDE, and the diagnostic picture changes only with longitudinal monitoring over several years. Estimates of the rates of diagnostic conversion vary widely, between 4.0% and >40%, depending in part on initial sample and the duration of follow-up (Angst, Sellaro, Stassen, & Gamma, 2005; Beesdo et al., 2009; Coryell et al., 1995; Goldberg, Harrow, & Whiteside, 2001), as summarized by Fiedorowicz et al. (2011). Subthreshold hypomanic symptoms are associated with the progression from unipolar to bipolar depression (Fiedorowicz et al., 2011), so a thorough inventory and history of symptomatology should be obtained as part of the diagnostic process. Other factors that are commonly associated with diagnostic conversion from MDD to bipolar disorder include early onset of affective symptoms, psychotic features, and a family history of mania. Although no diagnostic process can completely obviate the problem of diagnostic conversion, excluding individuals with a history of psychotic symptoms or a first-degree relative with bipolar disorder might further improve diagnostic confidence in the remaining subjects (in addition to improving safety, as noted below in the discussion of exclusion criteria).

### External verification

Mood is inherently subjective, and there are no objective biomarkers that can establish a clinical diagnosis of MDD. As such, depressive symptoms can easily be exaggerated or feigned, which can further complicate the question of accurate, reproducible diagnosis of potential subjects. This might be of particular concern in studies of psychedelic drugs, as the potential to use these drugs in a controlled clinical setting without the risk of legal sanctions may introduce an element of secondary gain. As such, it might be useful to favor patients who have already been in treatment for MDD, whether currently or in the past. This could be verified by clinical records, pharmacy records, or contact with the treating clinician. In addition, because hypothyroidism may manifest as depression, screening labs should include thyroid function studies to rule out this common organic etiology of depression. Patients with a known history of hypothyroidism who otherwise qualify for the study should be included only if they are euthyroid at the time of screening.

## Characteristics of the depressive episode

In addition to the diagnosis, the protocol should specify the duration of the episode, the severity of the symptoms, and the level of treatment resistance. If the patient population is too heterogeneous, a clinically significant antidepressant effect may be difficult to identify. This may be particularly true for patients whose depressive episode has atypical features, as the characteristic mood reactivity may obscure a treatment-specific effect. However, if the study population is defined too narrowly, the clinical value and generalizability will be compromised. The required duration of the MDE prior to the start of the study should be specified in the study protocol, as spontaneous remission is a potential confound in any study of MDD.

### Baseline depression severity

The placebo response in MDD is robust, even when baseline symptoms are severe (Elkin et al., 1995; Fournier et al., 2010; Wilcox et al., 1992). The placebo response varies inversely with depression severity – i.e., as baseline severity decreases, remission rates increase regardless of the treatment group. As such, higher baseline depression scores are associated with higher effect size in antidepressant clinical trials (Elkin et al., 1995; Khan, Leventhal, Khan, & Brown, 2002; Kirsch et al., 2008). To increase the likelihood of finding a true drug effect, depression clinical trials often restrict the study population to subjects with a Hamilton Depression Rating Scale (HAM-D) score of ≥23 (very severe depression), as defined by The American Psychiatric Association's *Handbook of Psychiatric Measures* (American Psychiatric Association, Task Force for the Handbook of Psychiatric Measures, & Rush, 2000). It is possible that such a restrictive cut-off might not be necessary in clinical studies of psilocybin for MDD. In their clinical trial of psilocybin in cancer-related anxiety and depression, the population studied by Ross et al. (2016) had very mild baseline depressive symptoms [baseline Beck

Depression Inventory (BDI) score for both groups combined: 16.04 ± 1.33], and the drug effect ($M \pm SE$ for BDI score at final post-crossover timepoint: 7.77 ± 1.42). In the study by Griffiths et al. (2016), participants' baseline depression severity was slightly higher (mean BDI = 17.77–18.40), the drug effect size was also large (Cohen's $d$ for BDI measured 5 weeks after a single high dose: 0.81). However, these studies should be interpreted with caution, as it is not clear whether the effect of psilocybin on depressive symptoms in cancer patients is generalizable to patients with MDD, particularly given the heterogeneity of both clinical populations.

## Prior treatment failures

In all cases, the number of failed medication trials in the current depressive episode should be reported, as should the total number of antidepressant trials over the course of the illness, if a patient has recurrent MDD. Since psilocybin may have a role as a rapid-acting antidepressant, it would be particularly useful to note whether the participant has a history of treatment failure with ketamine. The upstream pathophysiology of depressive symptoms in someone who previously responded to selective serotonin reuptake inhibitor (SSRI) monotherapy, for example, may be quite different from a depressive episode that is resistant to many more treatments and/or alternative modalities (e.g., electroconvulsive therapy; El-Hage, Leman, Camus, & Belzung, 2013). Limiting the level of treatment resistance in early trials may increase the consistency of clinical outcomes. This will be all the more important in trials that aim to study treatment-resistant (or treatment-refractory) depression (TRD) specifically. Although different studies define treatment resistance differently, it is generally understood to mean the failure to achieve remission despite at least two adequate antidepressant trials in the current depressive episode (Trevino, McClintock, McDonald Fischer, Vora, & Husain, 2014). Note that this definition does not take initial or post-treatment depression severity into account in a systematic way. As such, a clinically significant response (≥50% reduction in symptom severity) or partial response (25%–50% reduction) may have been achieved, even if the episode meets criteria for TRD (Berlim & Turecki, 2007; Nierenberg & DeCecco, 2001; Ruhé, van Rooijen, Spijker, Peeters, & Schene, 2012). Investigators should address all these issues when reporting results (Ruhé et al., 2012).

## Suicidality

A trial of any intervention for MDD should include longitudinal monitoring of suicidal ideation and behavior, and suicide risk should be taken into consideration when selecting participants. One large retrospective study of the general population found that psychedelic use is not associated with an increased risk of suicidality (Johansen & Krebs, 2015). On the contrary, some studies have found an association between psychedelic use and a reduced risk of suicidal ideation and behavior (Argento, Braschel, Walsh, Socias, & Shannon, 2018; Argento et al., 2017; Hendricks, Thorne, Clark, Coombs, & Johnson, 2015). Nevertheless, researchers studying psilocybin for MDD should have protocols in place to monitor for and mitigate an acutely elevated risk of suicide in study participants. We recommend that subjects be screened for suicidal ideation using the Columbia-Suicide Severity Rating Scale (CSSRS; Posner et al., 2011) at baseline and at each assessment point. A CSSRS score above a certain level may be exclusionary, but study-specific cut-offs are expected to vary depending on the level of care (e.g., a study of psychiatric inpatients might accept participants with a higher baseline CSSRS score than an outpatient study could safely accommodate). If a participant's CSSRS score were to increase over the course of the trial, careful clinical evaluation and reassessment of safety risk would be necessary. Escalation to a higher level of clinical care should be considered to maximize participant safety.

## Psychedelic clinical trials: Concurrent medications

After a diagnosis of MDD has been carried out, subjects are screened further to determine whether they meet additional inclusion/exclusion criteria. In most antidepressant drug trials, participants are not permitted to take other psychotropic medications during the study. Since many antidepressant medications directly modulate the serotonin system, including 5-HT$_{2A}$ receptors (Gómez-Gil et al., 2004; Meyer et al., 2001; Yamauchi, Miyara, Matsushima, & Imanishi, 2006), their continued use in a trial of a serotonergic hallucinogen might interact with the effect of the study drug in an unpredictable way. There is some evidence that chronic administration of different classes of antidepressant drugs have different effects on the intensity of the psychedelic experience, with patients taking a tricyclic antidepressant experiencing more and patients taking a monoamine oxidase inhibitor or an SSRI experiencing less potent subjective responses to LSD (Bonson, Buckholtz, & Murphy, 1996; Bonson & Murphy, 1996). We therefore recommend that efficacy studies require a washout period for antidepressants. If antidepressant efficacy is established for any classic hallucinogen, further studies may be warranted to determine whether they are safe and efficacious in the context of ongoing use of various classes of antidepressants, i.e., as augmenting agents.

There is a safety risk associated with an antidepressant washout period: participants might acutely decompensate if there was even a partial response to the prior treatment. The risk of this is relatively low, particularly in patients who meet criteria for TRD (Nugent, Iadarola, Miller, Luckenbaugh, & Zarate, 2016). However, subjects should be clinically reassessed during the washout period, including screening for suicidal ideation (e.g., with the CSSRS; Posner et al., 2011), to ensure that they do not require emergent psychiatric intervention. In addition, when reporting outcomes, the value used as the baseline comparator should be measured shortly (within 1 week) before the study drug is administered, rather than using the value obtained at the pre-washout screening.

## Other exclusion criteria

Given the acute psychotomimetic effects of psilocybin, patients at increased risk of psychosis have been excluded from prior studies demonstrating drug safety. We recommend doing the same in clinical trials of psilocybin for MDD.

This includes individuals with a personal history of psychosis (including MDD with psychotic features), comorbid borderline personality disorder, and/or first-degree relatives with schizophrenia, schizoaffective disorder, or bipolar disorder. Consistent with published protocols, we also recommend excluding patients with a history of coronary artery disease, cerebrovascular disease, or unstable medical conditions. This includes poorly controlled hypertension, as psilocybin is known to cause a transient increase in blood pressure.

Past psychedelic use is an important factor to consider when designing any trial in which psychedelic medications are administered, as prior experience with these drugs might increase the expectancy bias or decrease the psychological impact of a psychedelic experience. There are minimal empirical data to guide the decision of whether to exclude participants based on past use of psychedelics. However, most studies to date have imposed a cap, either on lifetime total uses, recent use, or both. We recommend excluding participants with any use of psychedelics (including psilocybin, MDMA, LSD, mescaline, DMT, and other similar hallucinogenic compounds) within the last year at a minimum. Excluding use for the past 5 years may be appropriate for some study designs. However, given the higher prevalence of recent psychedelic use in people in younger age groups (Krebs & Johansen, 2013), excluding patients with any use in the past 5 years might disproportionally reduce the number of eligible participants in those age groups.

## STUDY INTERVENTIONS

### Placebo selection and blinding

As discussed in the section on baseline symptom severity above, controlled trials of antidepressant medications consistently identify a robust placebo response. For this reason, uncontrolled studies of psilocybin's effect on mood should not be used as evidence of its antidepressant efficacy, as a suitable control group is essential for identifying a drug-specific effect (Barnby & Mehta, 2018). Given the powerful subjective effects of psychedelic drugs, there is no placebo that can fully preserve participant and clinician blinding in a controlled trial. This is frequently identified as a limitation of psychedelic studies described as double-blind. However, one should note that even in double-blind trials of established antidepressants, which have no psychedelic effects, a significant majority of both patients (78%) and clinicians (87%) accurately guess whether the patient was assigned to the treatment group or placebo (Rabkin et al., 1986). The fact that a fully blinded study is not possible for a psychedelic drug is therefore not a unique problem. Nevertheless, it is important to optimize study procedures to minimize the opportunity for bias.

### Placebo selection

The ideal control would be an active placebo, that is, a substance that lacks antidepressant efficacy but mimics some of the subjective effects of the study drug. Prior controlled studies of classic hallucinogens have used a variety of active placebos with variable subjective effects, including first-generation antihistamines, methylphenidate, and niacin. Very low doses of the study drug could also be a reasonable choice, and were used in the cancer study conducted by Griffiths et al. (2016). This might reduce the expectancy bias, as all participants will know they are receiving psilocybin. However, it might reduce between-group effect sizes, because it is possible that even very low doses of psychedelics could have significant antidepressant effects, as the growing interest in psychedelic "microdosing" (Fadiman, 2011) suggests.

### Independent raters

In addition to the use of an active placebo, the most important way to further reduce expectation bias is for independent raters to conduct the assessments used as primary outcomes (Khan & Brown, 2015). The raters should not in any way be a part of the treatment team for that particular subject, and the treatment team should not discuss details of the dosing sessions with the independent raters. The participant should have no other interaction with the rater outside of the rating sessions, and they should be explicitly instructed not to talk with the rater about the experience they had during the dosing session. The independent rater(s) should receive formal training, as this affects the interrater reliability of the scales (Cusin, Yang, Yeung, & Fava, 2009).

### Dose selection

The dosing regimen comprises the amount of active drug, including the choice of a weight-based or flat-dosing schedule, and the number and timing of medication sessions. In the two recent trials for treatment of cancer-related anxiety and/or depression, participants received a single dose of psilocybin (Griffiths et al., 2016; Ross et al., 2016). By contrast, open-label pilot studies of psilocybin for smoking cessation (Johnson, Garcia-Romeu, Cosimano, & Griffiths, 2014; Johnson, Garcia-Romeu, & Griffiths, 2017), alcohol use disorder (Bogenschutz et al., 2015), and TRD (Carhart-Harris et al., 2016, 2017) have used two or more dosing sessions. Each of these trials reported participant benefits for the various indications under investigation. With respect to the dose itself, dose-response data comparing 30 mg/70 kg (0.43 mg/kg) and 20 mg/70 kg (0.29 mg/kg) in healthy volunteers suggest that psychological adverse events such as anxiety and paranoia may be more likely following the higher dose (Griffiths et al., 2011). A recent report of psilocin clearance and subjective effects of orally administered psilocybin suggests that a fixed dose of 25 mg psilocybin is likely to approximate circulating levels of psilocin following a weight-adjusted dose of 0.3 mg/kg (21 mg/70 kg; Brown et al., 2017). Flat dosing is logistically appealing, as it obviates the need for patient-specific pharmacy compounding and allows randomization to occur immediately before the medication session.

The protocol in our ongoing double-blind trial of psilocybin-assisted psychotherapy for alcohol use disorder at the NYU School of Medicine (ClinicalTrials.gov Identifier: NCT02061293) includes 2–3 medication sessions

with flexible dosing embedded in an extensive psychosocial framework. Participants randomized to the psilocybin group receive an initial dose of 0.36 mg/kg (25 mg/70 kg). Based on the safety data and scores on the acute hallucinogen rating scales, the dose can be increased to 0.43 mg/kg (30 mg/70 kg) or 0.57 mg/kg (40 mg/70 kg) in subsequent sessions. The decision to incorporate flexible, stepwise dosing was made to maximize the potential for each participant to have a strong subjective experience while minimizing the risk of adverse psychiatric or medical events. This is because prior studies of drinking outcomes suggested that the intensity of the acute experience was correlated with improved drinking outcomes. However, it is not clear whether the same would be true in MDD. In their open-label pilot study of psilocybin for TRD, Carhart-Harris et al. (2016, 2017) measured clinical outcomes after subjects received two doses of psilocybin, first low (10 mg) and then high (25 mg), 1 week apart. In the initial report, even after the low dose, depression severity (as measured by HAM-D scores) fell dramatically after 1 week, from $21.4 \pm 4.5$ to $10.7 \pm 7.8$. The second medication session brought scores an additional 3.3 points lower 1 week later (HAM-D score $7.4 \pm 6.9$). It is not clear whether this further reduction was due to the fact that the second dose was higher than the first; it is possible that a second medication session with the 10 mg dose would have caused the same reduction in depressive symptoms. If early randomized, controlled trials indicate antidepressant efficacy, dose-response studies will be needed to explore these questions further.

### Psychological support

In psychedelic clinical research, the standard practice is for drug-administration sessions to be conducted by a mixed-gender team of two therapists, interventionists, or guides. The therapists provide at least several hours of preparatory sessions in addition to any indication-specific therapy. They perform careful monitoring and provide non-directive support during the medication session itself. They also meet with the participant (often several times) after the medication session to allow the opportunity to process or integrate the experience, and to monitor for any negative sequelae of the sessions (Johnson, Richards, & Griffiths, 2008). Although the components of this psychosocial platform are based on tradition and experience, studies conducted with such a supportive framework have successfully minimized the occurrence of adverse psychological events during and after the dosing session. It may also be integral to the therapeutic effect, particularly if that effect depends on psychological experiences that require context and integration. The specific number and content of preparatory and debriefing sessions will likely vary across studies, as will the dosing schedules. The qualifications, training, and gender combinations of the therapists will also likely vary across studies. Although an attempt to unify approaches at this point would be premature, some limits on heterogeneity are necessary, in order to highlight the drug-specific effect and maximize the extent to which outcomes can be compared across studies. For example, a depression-focused psychotherapy platform might obscure the drug-specific effect, and would not be preferred in a drug

efficacy study. However, combining psilocybin with a depression-focused psychotherapy platform might be highly relevant in the treatment optimization studies that would follow if efficacy studies have positive results. It is unclear whether the antidepressant effect of psilocybin might allow for its use as monotherapy, or whether an optimal effect can be achieved only when it is used as an adjunct to psychotherapy. In their open-label study of psilocybin for TRD, Carhart-Harris et al. (2016, 2017) provided comparatively little psychological preparation and debriefing. Despite a robust short-term antidepressant effect, by 3 months, 7 of 12 patients had relapsed, reporting mild-moderate or severe depressive symptoms. Different treatment protocols (e.g., a higher number of debriefing/integration sessions) might be associated with significant differences in short- and long-term clinical effects, which could guide future studies designed to optimize the treatment.

## SAFETY MONITORING

Human studies to date have confirmed that psilocybin has a relatively benign safety profile when administered in a clinical setting (Bogenschutz et al., 2015; Carhart-Harris et al., 2016; Griffiths et al., 2011, 2016; Grob et al., 2011; Ross et al., 2016). However, the importance of setting should not be underestimated, and it is not at all clear that recreational use of psilocybin would have an equally favorable safety profile. To minimize the risk and maximize the potential benefit to the participant, a number of intrinsic and extrinsic factors should be considered (Johnson et al., 2008). With respect to intrinsic factors, the exclusion criteria should rule out individuals with an elevated risk of psychosis (e.g., prior history of psychotic episode, personal or first-degree relative with a primary psychotic illness, or bipolar disorder). Participants must be adequately prepared prior to the medication session, and post-medication debriefing sessions should be provided to assist the participant in integrating the drug experience. With respect to extrinsic factors, the drug should be administered in a relaxing and comfortable environment. During the dosing session, the participants should be continuously monitored by individuals with whom they have established rapport. There should be safety protocols in place for the management of rare but potentially serious adverse reactions – including hypertensive urgency, intractable anxiety, or acute psychosis – that might require urgent or emergent intervention. In addition, participants should be required to remain on-site for a minimum of 6 hr, preferably longer, after the medication is administered. They should be monitored until any perceptual changes have fully resolved. Participants should not be permitted to drive immediately after the session, and ideally will be accompanied home.

## TRIAL DURATION

In antidepressant drug trials, the placebo effect occurs early and is variable, whereas a stable, drug-specific (rather than placebo) effect is more likely to be measured after the first 2 weeks of treatment (Quitkin, 1992; Quitkin, Rabkin,

Ross, & Stewart, 1984). A meta-analysis of data from 104 antidepressant trials indicated that the minimum adequate length of a drug trial is 4 weeks (Tedeschini, Fava, & Papakostas, 2011). We recommend that participants be followed at least 6 weeks after a single treatment, to determine whether there are any durable antidepressant effects that last beyond the "afterglow" period, which can last up to a month after the dosing session (Majić, Schmidt, & Gallinat, 2015). However, as the trial duration increases, the possibility of spontaneous remission rises, so a drug-specific antidepressant effect may wash out. The antidepressant effect should also be measured soon after the dosing session, consistent with clinical trials of ketamine, in which an antidepressant effect is observed within 24 hr of, and up to 7 days after, a single treatment (Berman et al., 2000; Zarate et al., 2006). The data in prior studies of psilocybin suggest that the antidepressant effect of a single dose may also be rapid, and may even last longer than ketamine. To further explore the possibility, we recommend that the primary outcome be measured 1 week after the dosing session. Symptom-specific analyses should be considered for assessments that take place less than 24 hr after drug administration, to avoid scoring symptoms that require at least 24 hr to accurately evaluate, such as appetite or sleep.

### Anticipating attrition

In a clinical trial of a psychedelic drug, a participant who undergoes a perceptually and psychologically uneventful dosing session is likely to assume that they have received the control drug. In some cases, this might cause disappointment that could limit the participant's motivation to continue with follow-up assessments. One way to reduce the risk of attrition in those cases is to introduce an optional extension phase at the end of the trial, where participants are offered a dose of the active drug in an unblinded fashion. This unblinded active medication session mitigates potential disappointment after a presumed placebo dosing session during the main trial, incentivizes continued participation, and allows additional safety data to be collected. In studies of MDD, inclusion of an unblinded medication session has the additional benefit of minimizing safety concerns associated with disappointment in depressed patients following an uneventful medication session.

The major limitation of offering an unblinded session is that long-term follow-up and between-group comparisons become complex to evaluate once both groups have received active medication. For this reason, it may not be preferred in a major clinical trial of psilocybin for MDD. Nonetheless, most ongoing psychedelic trials offer an unblinded session to participants for the reasons listed above. Our ongoing phase-II double-blind trial of psilocybin for alcohol dependence includes an unblinded extension phase at the end of the double-blind follow-up phase that is available to all participants, regardless of their group assignment. An alternative option is to offer an active medication session only to participants who were randomized to receive placebo treatment. This can be done on a participant-by-participant basis by breaking the blind after each participant completes the double-blind phase of the trial, or (more conservatively) by waiting until data collection is complete and the blind has been broken for all participants. The latter option is the design that was chosen for the recently published MAPS phase-II studies of MDMA for post-traumatic stress disorder (Mithoefer et al., 2019). A crossover design in which participants receive active study medication in only one of two blinded medication sessions was another possibility, and was the design of choice for the trials of cancer-associated anxiety and depression completed by Griffiths and Ross and colleagues (Griffiths et al., 2016; Ross et al., 2016). Regardless of the design chosen, we recommend that participants be required to complete all therapy and assessment visits to be eligible for an optional unblinded active medication session.

## OUTCOME MEASURES

The goal of any treatment is remission, i.e., a total HAM-D score of $\leq 7$ or a Montgomery–Asberg Depression Rating Scale (MADRS) score of $\leq 10$. The rate of *response*, defined as a decrease of 50% or more from baseline symptom rating score, should also be reported. In a clinical trial of psilocybin for MDD, the acute subjective effects of the drug should be approached as a side effect, but also investigated as a possible mediator of clinical response.

### Clinician-rated scales

Historically, clinician-rated depression scores are the gold standard primary outcome measure in antidepressant efficacy trials. The three clinician-rated scales most commonly used in RCTs are the HAM-D, the MADRS, and the 16-item clinician-rated version of the Quick Inventory of Depression Scale (QIDS-C$_{16}$). The HAM-D is the oldest measure, and has been the most widely used in clinical trials over the past several decades (Brunoni, Lopes, Kaptchuk, & Fregni, 2009; Fournier et al., 2010; Kirsch et al., 2008), so it will be used in the discussion of severity measurement below. However, for a number of reasons, it has attracted considerable criticism (Bagby, Ryder, Schuller, & Marshall, 2004; Huijbrechts, Haffmans, Jonker, van Dijke, & Hoencamp, 1999; Uher et al., 2008). Chief among these is the fact that multiple symptoms included in the DSM-IV and DSM-V diagnostic criteria for MDD are not scored on the HAM-D (e.g., poor concentration), whereas symptoms such as hypochondriasis and loss of insight, which are not among the diagnostic criteria for MDD, are scored. For this and other reasons, the MADRS is increasingly preferred over the HAM-D in clinical trials of MDD (Carmody et al., 2006; Huijbrechts et al., 1999; Uher et al., 2008). One major limitation of the MADRS is its failure to measure reverse neurovegetative symptoms (hypersomnia and hyperphagia). An acceptable alternative is the QIDS-C$_{16}$ (Rush et al., 2003; Trivedi et al., 2004). It is used in the STAR*D depression trial (Rush et al., 2004), measures only the nine criterion domains of MDD, has good internal validity, and is freely available.

### Subject-rated scales

Although clinician-rated scales are the gold standard in efficacy trials and should be the primary clinical outcome,

a subject-rated (self-report) scale should be reported as a secondary outcome. The use of self-report (e.g., the BDI; Beck, Steer, Ball, & Ranieri, 1996) has the advantage of circumventing the clinician's tendency to selectively overestimate severity at baseline compared to later time points (Tohen, Bowden, Nierenberg, & Geddes, 2015). The concordance between clinician-rated and self-reported scores is moderate (John Mann et al., 2016; Zimmerman, Walsh, Friedman, Boerescu, & Attiullah, 2017). The consensus is that clinician-rated and self-reported scales provide complementary measures of depression severity and should both be assessed in clinical trials (Dunlop, McCabe, Eudicone, Sheehan, & Baker, 2014; Greenberg, Bornstein, Greenberg, & Fisher, 1992; Rane et al., 2010; Uher et al., 2012). Although the BDI is the gold standard, it is copyrighted, which limits its use. The QIDS-SR$_{16}$ is an acceptable non-proprietary alternative (Tada et al., 2014; Trivedi et al., 2004).

### Additional analyses

Depression rating scales provide information regarding the overall severity of a depressive episode and its general responsiveness to treatment. With rigorous statistical controls, scores on these scales can be analyzed further to examine the effect of the treatment on individual symptoms (e.g., sleep and suicidality) or symptom clusters (e.g., cognitive disturbances and social interactions). Carefully controlled analyses might generate hypotheses about specific dimensions of illness that could be targeted in future studies of therapeutic mechanisms.

## CONCLUSIONS

There is an urgent need to develop novel treatments for MDD, and there are many reasons to hypothesize that psychedelic treatment might meet this need. However, such an effect has not been demonstrated by the standards of contemporary clinical research, and researchers in this field will be well served by maintaining a skeptical approach when designing, conducting, and disseminating their studies. Given the subjective quality and wide variability of depressive symptomatology, as well as the inherent safety risks and the high placebo response in antidepressant trials, great care should be paid when designing a protocol to study a putative treatment for MDD. We have discussed some of the factors, including diagnostic accuracy, symptom measurement, concurrent treatment, and rater blinding, as well as the severity, duration, and treatment resistance of the current episode, and practical considerations for administration of classic hallucinogens to treat MDD that should be carefully deliberated when designing a protocol. We have focused on psilocybin, although other psychedelics are being studied, including a recently published trial of ayahuasca for MDD (Palhano-Fontes et al., 2018). An additional factor that has not been discussed in detail is the content (including the theoretical framework) of the supportive framework, including any preparatory and debriefing psychotherapy sessions, which are expected to vary substantially across studies. When protocols are published,

researchers should provide a full and transparent account of why each factor was chosen. This will ensure that the highest standards for contemporary trial design are met. Rigorous research protocols will increase the likelihood of producing meaningful, reproducible outcomes. It will also hasten the broader acceptance of psychedelic drug research in the scientific community at large.

## REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (DSM–IV)* (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)* (5th ed.). Arlington, VA: American Psychiatric Publishing.

American Psychiatric Association, Task Force for the Handbook of Psychiatric Measures, & Rush, A. J. (2000). *Handbook of psychiatric measures* (1st ed.). Washington, DC: American Psychiatric Association.

Angst, J., Sellaro, R., Stassen, H. H., & Gamma, A. (2005). Diagnostic conversion from depression to bipolar disorders: Results of a long-term prospective study of hospital admissions. *Journal of Affective Disorders, 84*(2–3), 149–157. doi:10.1016/S0165-0327(03)00195-2

Argento, E., Braschel, M., Walsh, Z., Socias, M. E., & Shannon, K. (2018). The moderating effect of psychedelics on the prospective relationship between prescription opioid use and suicide risk among marginalized women. *Journal of Psychopharmacology, 32*(12), 1385–1391. doi:10.1177/0269881118798610

Argento, E., Strathdee, S. A., Tupper, K., Braschel, M., Wood, E., & Shannon, K. (2017). Does psychedelic drug use reduce risk of suicidality? Evidence from a longitudinal community-based cohort of marginalised women in a Canadian setting. *BMJ Open, 7*(9), e016025. doi:10.1136/bmjopen-2017-016025

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). The Hamilton Depression Rating Scale: Has the gold standard become a lead weight? *The American Journal of Psychiatry, 161*(12), 2163–2177. doi:10.1176/appi.ajp.161.12.2163

Barnby, J. M., & Mehta, M. A. (2018). Psilocybin and mental health – Don't lose control. *Frontiers in Psychiatry, 9,* 293. doi:10.3389/fpsyt.2018.00293

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. (1996). Comparison of Beck Depression Inventories – IA and -II in psychiatric outpatients. *Journal of Personality Assessment, 67*(3), 588–597. doi:10.1207/s15327752jpa6703_13

Beesdo, K., Hofler, M., Leibenluft, E., Lieb, R., Bauer, M., & Pfennig, A. (2009). Mood episodes and mood disorders:

Patterns of incidence and conversion in the first three decades of life. *Bipolar Disorders, 11*(6), 637–649. doi:10.1111/j.1399-5618.2009.00738.x

Berlim, M. T., & Turecki, G. (2007). What is the meaning of treatment resistant/refractory major depression (TRD)? A systematic review of current randomized trials. *European Neuropsychopharmacology, 17*(11), 696–707. doi:10.1016/j.euroneuro.2007.03.009

Berman, R. M., Cappiello, A., Anand, A., Oren, D. A., Heninger, G. R., Charney, D. S., & Krystal, J. H. (2000). Antidepressant effects of ketamine in depressed patients. *Biological Psychiatry, 47*(4), 351–354. doi:10.1016/s0006-3223(99)00230-9

Bogenschutz, M. P., Forcehimes, A. A., Pommy, J. A., Wilcox, C. E., Barbosa, P. C., & Strassman, R. J. (2015). Psilocybin-assisted treatment for alcohol dependence: A proof-of-concept study. *Journal of Psychopharmacology, 29*(3), 289–299. doi:10.1177/0269881114565144

Bonson, K. R., Buckholtz, J. W., & Murphy, D. L. (1996). Chronic administration of serotonergic antidepressants attenuates the subjective effects of LSD in humans. *Neuropsychopharmacology, 14*(6), 425–436. doi:10.1016/0893-133X(95)00145-4

Bonson, K. R., & Murphy, D. L. (1996). Alterations in responses to LSD in humans associated with chronic administration of tricyclic antidepressants, monoamine oxidase inhibitors or lithium. *Behavioural Brain Research, 73*(1–2), 229–233. doi:10.1016/0166-4328(96)00102-7

Brown, R. T., Nicholas, C. R., Cozzi, N. V., Gassman, M. C., Cooper, K. M., Muller, D., Thomas, C. D., Hetzel, S. J., Henriquez, K. M., Ribaudo, A. S., & Hutson, P. R. (2017). Pharmacokinetics of escalating doses of oral psilocybin in healthy adults. *Clinical Pharmacokinetics, 56*(12), 1543–1554. doi:10.1007/s40262-017-0540-6

Brunoni, A. R., Lopes, M., Kaptchuk, T. J., & Fregni, F. (2009). Placebo response of non-pharmacological and pharmacological trials in major depression: A systematic review and meta-analysis. *PLoS One, 4*(3), e4824. doi:10.1371/journal.pone.0004824

Carhart-Harris, R. L., Bolstridge, M., Day, C. M. J., Rucker, J., Watts, R., Erritzoe, D. E., Kaelen, M., Giribaldi, B., Bloomfield, M., Pilling, S., Rickard, J. A., Forbes, B., Feilding, A., Taylor, D., Curran, H. V., & Nutt, D. J. (2017). Psilocybin with psychological support for treatment-resistant depression: Six-month follow-up. *Psychopharmacology (Berl), 235*(2), 399–408. doi:10.1007/s00213-017-4771-x

Carhart-Harris, R. L., Bolstridge, M., Rucker, J., Day, C. M., Erritzoe, D., Kaelen, M., Bloomfield, M., Rickard, J. A., Forbes, B., Feilding, A., Taylor, D., Pilling, S., Curran, V. H., & Nutt, D. J. (2016). Psilocybin with psychological support for treatment-resistant depression: An open-label feasibility study. *Lancet Psychiatry, 3*(7), 619–627. doi:10.1016/S2215-0366(16)30065-7

Carmody, T. J., Rush, A. J., Bernstein, I., Warden, D., Brannan, S., Burnham, D., Woo, A., & Trivedi, M. H. (2006). The Montgomery Asberg and the Hamilton ratings of depression: A comparison of measures. *European Neuropsychopharmacology, 16*(8), 601–611. doi:10.1016/j.euroneuro.2006.04.008

Coryell, W., Endicott, J., Maser, J. D., Keller, M. B., Leon, A. C., & Akiskal, H. S. (1995). Long-term stability of polarity distinctions in the affective disorders. *The American Journal of Psychiatry, 152*(3), 385–390. doi:10.1176/ajp.152.3.385

Cusin, C., Yang, H., Yeung, A., & Fava, M. (2009). Rating scales for depression. In L. Baer & M. A. Blais (Eds.), *Handbook of clinical rating scales and assessment in psychiatry and mental health. Current clinical psychiatry*. Totowa, NJ: Humana Press.

Dunlop, B. W., McCabe, B., Eudicone, J. M., Sheehan, J. J., & Baker, R. A. (2014). How well do clinicians and patients agree on depression treatment outcomes? Implications for personalized medicine. *Human Psychopharmacology, 29*(6), 528–536. doi:10.1002/hup.2428

El-Hage, W., Leman, S., Camus, V., & Belzung, C. (2013). Mechanisms of antidepressant resistance. *Frontiers in Pharmacology, 4,* 146. doi:10.3389/fphar.2013.00146

Elkin, I., Gibbons, R. D., Shea, M. T., Sotsky, S. M., Watkins, J. T., Pilkonis, P. A., & Hedeker, D. (1995). Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 63*(5), 841–847. doi:10.1037//0022-006x.63.5.841

Fadiman, J. (2011). *The psychedelic explorer's guide: Safe, therapeutic, and sacred journeys*. Rochester, VT: Park Street Press.

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J., Vos, T., & Whiteford, H. A. (2013). Burden of depressive disorders by country, sex, age, and year: Findings from the global burden of disease study 2010. *PLoS Medicine, 10*(11), e1001547. doi:10.1371/journal.pmed.1001547

Fiedorowicz, J. G., Endicott, J., Leon, A. C., Solomon, D. A., Keller, M. B., & Coryell, W. H. (2011). Subthreshold hypomanic symptoms in progression from unipolar major depression to bipolar disorder. *The American Journal of Psychiatry, 168*(1), 40–48. doi:10.1176/appi.ajp.2010.10030328

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C., & Fawcett, J. (2010). Antidepressant drug effects and depression severity: A patient-level meta-analysis. *JAMA, 303*(1), 47–53. doi:10.1001/jama.2009.1943

Freedman, R., Lewis, D. A., Michels, R., Pine, D. S., Schultz, S. K., Tamminga, C. A., Gabbard, G. O., Gau, S. S., Javitt, D. C., Oquendo, M. A., Shrout, P. E., Vieta, E., & Yager, J. (2013). The initial field trials of DSM-5: New blooms and old thorns. *The American Journal of Psychiatry, 170*(1), 1–5. doi:10.1176/appi.ajp.2012.12091189

Goldberg, J. F., Harrow, M., & Whiteside, J. E. (2001). Risk for bipolar illness in patients initially hospitalized for unipolar depression. *The American Journal of Psychiatry, 158*(8), 1265–1270. doi:10.1176/appi.ajp.158.8.1265

Gómez-Gil, E., Gastó, C., Carretero, M., Díaz-Ricart, M., Salamero, M., Navinés, R., & Escolar, G. (2004). Decrease of the platelet 5-HT2A receptor function by long-term imipramine treatment in endogenous depression. *Human Psychopharmacology, 19*(4), 251–258. doi:10.1002/hup.583

Greenberg, R. P., Bornstein, R. F., Greenberg, M. D., & Fisher, S. (1992). A meta-analysis of antidepressant outcome under "blinder" conditions. *Journal of Consulting and Clinical Psychology, 60*(5), 664–669; discussion 670–667. doi:10.1037//0022-006x.60.5.664

Griffiths, R. R., Johnson, M. W., Carducci, M. A., Umbricht, A., Richards, W. A., Richards, B. D., Cosimano, M. P., & Klinedinst, M. A. (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. doi:10.1177/0269881116675513

Griffiths, R. R., Johnson, M. W., Richards, W. A., Richards, B. D., McCann, U., & Jesse, R. (2011). Psilocybin occasioned mystical-type experiences: Immediate and persisting dose-related effects. *Psychopharmacology (Berl), 218*(4), 649–665. doi:10.1007/s00213-011-2358-5

Grob, C. S., Danforth, A. L., Chopra, G. S., Hagerty, M., McKay, C. R., Halberstadt, A. L., & Greer, G. R. (2011). Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. *Archives of General Psychiatry, 68*(1), 71–78. doi:10.1001/archgenpsychiatry.2010.116

Hendricks, P. S., Thorne, C. B., Clark, C. B., Coombs, D. W., & Johnson, M. W. (2015). Classic psychedelic use is associated with reduced psychological distress and suicidality in the United States adult population. *Journal of Psychopharmacology, 29*(3), 280–288. doi:10.1177/0269881114565653

Huijbrechts, I. P., Haffmans, P. M., Jonker, K., van Dijke, A., & Hoencamp, E. (1999). A comparison of the 'Hamilton Rating Scale for Depression' and the 'Montgomery-Asberg Depression Rating Scale'. *Acta Neuropsychiatrica, 11*(1), 34–37. doi:10.1017/S0924270800036358

Johansen, P. O., & Krebs, T. S. (2015). Psychedelics not linked to mental health problems or suicidal behavior: A population study. *Journal of Psychopharmacology, 29*(3), 270–279. doi:10.1177/0269881114568039

John Mann, J., Ellis, S. P., Currier, D., Zelazny, J., Birmaher, B., Oquendo, M. A., Stanley, B., Melhem, N., Burke, A. K., & Brent, D. A. (2016). Self-rated depression severity relative to clinician-rated depression severity: Trait stability and potential role in familial transmission of suicidal behavior. *Archives of Suicide Research, 20*(3), 412–425. doi:10.1080/13811118.2015.1033504

Johnson, M., Richards, W., & Griffiths, R. (2008). Human hallucinogen research: Guidelines for safety. *Journal of Psychopharmacology, 22*(6), 603–620. doi:10.1177/0269881108093587

Johnson, M. W., Garcia-Romeu, A., Cosimano, M. P., & Griffiths, R. R. (2014). Pilot study of the 5-HT2AR agonist psilocybin in the treatment of tobacco addiction. *Journal of Psychopharmacology, 28*(11), 983–992. doi:10.1177/0269881114548296

Johnson, M. W., Garcia-Romeu, A., & Griffiths, R. R. (2017). Long-term follow-up of psilocybin-facilitated smoking cessation. *The American Journal of Drug and Alcohol Abuse, 43*(1), 55–60. doi:10.3109/00952990.2016.1170135

Khan, A., & Brown, W. A. (2015). Antidepressants versus placebo in major depression: An overview. *World Psychiatry, 14*(3), 294–300. doi:10.1002/wps.20241

Khan, A., Leventhal, R. M., Khan, S. R., & Brown, W. A. (2002). Severity of depression and response to antidepressants and placebo: An analysis of the Food and Drug Administration database. *Journal of Clinical Psychopharmacology, 22*(1), 40–45. doi:10.1097/00004714-200202000-00007

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine, 5*(2), e45. doi:10.1371/journal.pmed.0050045

Krebs, T. S., & Johansen, P. O. (2013). Over 30 million psychedelic users in the United States. *F1000Research, 2,* 98. doi:10.12688/f1000research.2-98.v1

Lamers, F., Cui, L., Hickie, I. B., Roca, C., Machado-Vieira, R., Zarate, C. A., & Merikangas, K. R. (2016). Familial aggregation and heritability of the melancholic and atypical subtypes of depression. *Journal of Affective Disorders, 204,* 241–246. doi:10.1016/j.jad.2016.06.040

Lobbestael, J., Leurgans, M., & Arntz, A. (2011). Inter-rater reliability of the structured clinical interview for DSM-IV Axis I disorders (SCID I) and Axis II disorders (SCID II). *Clinical Psychology & Psychotherapy, 18*(1), 75–79. doi:10.1002/cpp.693

Majić, T., Schmidt, T. T., & Gallinat, J. (2015). Peak experiences and the afterglow phenomenon: When and how do therapeutic effects of hallucinogens depend on psychedelic experiences? *Journal of Psychopharmacology, 29*(3), 241–253. doi:10.1177/0269881114568040

Meyer, J. H., Kapur, S., Eisfeld, B., Brown, G. M., Houle, S., DaSilva, J., Wilson, A. A., Rafi-Tari, S., Mayberg, H. S., & Kennedy, S. H. (2001). The effect of paroxetine on 5-HT(2A) receptors in depression: An [(18)F]setoperone PET imaging study. *The American Journal of Psychiatry, 158*(1), 78–85. doi:10.1176/appi.ajp.158.1.78

Mithoefer, M. C., Feduccia, A. A., Jerome, L., Mithoefer, A., Wagner, M., Walsh, Z., Hamilton, S., Yazar-Klosinski, B., Emerson, A., & Doblin, R. (2019). MDMA-assisted psychotherapy for treatment of PTSD: Study design and rationale for phase 3 trials based on pooled analysis of six phase 2 randomized controlled trials. *Psychopharmacology, 236*(9), 2735–2745. doi:10.1007/s00213-019-05249-5

Nierenberg, A. A., & DeCecco, L. M. (2001). Definitions of antidepressant treatment response, remission, nonresponse, partial response, and other relevant outcomes: A focus on treatment-resistant depression. *The Journal of Clinical Psychiatry, 62*(Suppl. 16), 5–9.

Nugent, A. C., Iadarola, N. D., Miller, F. G., Luckenbaugh, D. A., & Zarate, C. A. (2016). Safety of research into severe and treatment-resistant mood disorders: Analysis of outcome data from 12 years of clinical trials at the US National Institute of Mental Health. *Lancet Psychiatry, 3*(5), 436–442. doi:10.1016/S2215-0366(16)00006-7

Palhano-Fontes, F., Barreto, D., Onias, H., Andrade, K. C., Novaes, M. M., Pessoa, J. A., Mota-Rolim, S. A., Osório, F. L., Sanches, R., Dos Santos, R. G., Tófoli, L. F., de Oliveira Silveira, G., Yonamine, M., Riba, J., Santos, F. R., Silva-Junior, A. A., Alchieri, J. C., Galvão-Coelho, N. L., Lobão-Soares, B., Hallak, J. E. C., Arcoverde, E., Maia-de-Oliveira, J. P., & Araújo, D. B. (2018). Rapid antidepressant effects of the psychedelic ayahuasca in treatment-resistant depression: A randomized placebo-controlled trial. *Psychological Medicine, 49*(4), 655–663. doi:10.1017/s0033291718001356

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., & Mann, J. J. (2011). The Columbia–Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *The American Journal of Psychiatry, 168*(12), 1266–1277. doi:10.1176/appi.ajp.2011.10111704

Power, R. A., Tansey, K. E., Buttenschøn, H. N., Cohen-Woods, S., Bigdeli, T., Hall, L. S., Kutalik, Z., Lee, S. H., Ripke, S., Steinberg, S., Teumer, A., Viktorin, A., Wray, N. R., Arolt, V., Baune, B. T., Boomsma, D. I., Børglum, A. D., Byrne, E. M., Castelao, E., Craddock, N., Craig, I. W., Dannlowski, U., Deary, I. J., Degenhardt, F., Forstner, A. J., Gordon, S. D.,

Grabe, H. J., Grove, J., Hamilton, S. P., Hayward, C., Heath, A. C., Hocking, L. J., Homuth, G., Hottenga, J. J., Kloiber, S., Krogh, J., Landén, M., Lang, M., Levinson, D. F., Lichtenstein, P., Lucae, S., MacIntyre, D. J., Madden, P., Magnusson, P. K. E., Martin, N. G., McIntosh, A. M., Middeldorp, C. M., Milaneschi, Y., Montgomery, G. W., Mors, O., Müller-Myhsok, B., Nyholt, D. R., Oskarsson, H., Owen, M. J., Padmanabhan, S., Penninx, B. W. J. H., Pergadia, M. L., Porteous, D. J., Potash, J. B., Preisig, M., Rivera, M., Shi, J., Shyn, S. I., Sigurdsson, E., Smit, J. H., Smith, B. H., Stefansson, H., Stefansson, K., Strohmaier, J., Sullivan, P. F., Thomson, P., Thorgeirsson, T. E., Van der Auwera, S., Weissman, M. M., CONVERGE Consortium, CARDIoGRAM Consortium, GERAD Consortium, Breen, G., & Lewis, C. M. (2016). Genome-wide association for major depression through age at onset stratification: Major depressive disorder working group of the psychiatric genomics consortium. *Biological Psychiatry, 81*(4), 325–335. doi:10.1016/j.biopsych.2016.05.010

Quitkin, F. M. (1992). Methodology of measuring the efficacy of antidepressants. *Psychopharmacology (Berl), 106*(Suppl. 1), S87–S89. doi:10.1007/BF02246244

Quitkin, F. M., Rabkin, J. G., Ross, D., & Stewart, J. W. (1984). Identification of true drug response to antidepressants. Use of pattern analysis. *Archives of General Psychiatry, 41*(8), 782–786. doi:10.1001/archpsyc.1984.01790190056007

Rabkin, J. G., Markowitz, J. S., Stewart, J., McGrath, P., Harrison, W., Quitkin, F. M., & Klein, D. F. (1986). How blind is blind? Assessment of patient and doctor medication guesses in a placebo-controlled trial of imipramine and phenelzine. *Psychiatry Research, 19*(1), 75–86. doi:10.1016/0165-1781(86)90094-6

Rane, L. J., Fekadu, A., Wooderson, S., Poon, L., Markopoulou, K., & Cleare, A. J. (2010). Discrepancy between subjective and objective severity in treatment-resistant depression: Prediction of treatment outcome. *Journal of Psychiatric Research, 44*(15), 1082–1087. doi:10.1016/j.jpsychires.2010.03.020

Ross, S., Bossis, A., Guss, J., Agin-Liebes, G., Malone, T., Cohen, B., Mennenga, S. E., Belser, A., Kalliontzi, K., Babb, J., Su, Z., Corby, P., & Schmidt, B. L. (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. doi:10.1177/0269881116675512

Ruhé, H. G., van Rooijen, G., Spijker, J., Peeters, F. P., & Schene, A. H. (2012). Staging methods for treatment resistant depression. A systematic review. *Journal of Affective Disorders, 137*(1–3), 35–45. doi:10.1016/j.jad.2011.02.020

Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., Thase, M. E., Nierenberg, A. A., MQuitkin, F., Kashner, T. M., Kupfer, D. J., Rosenbaum, J. F., Alpert, J., Stewart, J. W., McGrath, P. J., Biggs, M. M., Shores-Wilson, K., Lebowitz, B. D., Ritz, L., Niederehe, G., & STAR*D Investigators Group. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design. *Controlled Clinical Trials, 25*(1), 119–142. doi:10.1016/S0197-2456(03)00112-0

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573–583. doi:10.1016/s0006-3223(02)01866-8

Schatzberg, A. F. (2019). Scientific issues relevant to improving the diagnosis, risk assessment, and treatment of major depression. *American Journal of Psychiatry, 176*(5), 342–347. doi:10.1176/appi.ajp.2019.19030273

Sedlackova, Z., Prasko, J., Latalova, K., Kamaradova, D., Ociskova, M., Grambal, A., Sigmundova, Z., Kasalova, P., & Cakirpaloglu, S. (2015). Psychosocial aspects of resistance in complex treatment of depressive disorder. *Neuroendocrinology Letters, 36*(4), 354–362.

Sidor, M. M., & Macqueen, G. M. (2011). Antidepressants for the acute treatment of bipolar depression: A systematic review and meta-analysis. *The Journal of Clinical Psychiatry, 72*(2), 156–167. doi:10.4088/JCP.09r05385gre

Tada, M., Uchida, H., Suzuki, T., Abe, T., Pollock, B. G., & Mimura, M. (2014). Baseline difference between patients' and clinicians' rated illness severity scores and subsequent outcomes in major depressive disorder: Analysis of the sequenced treatment alternatives to relieve depression data. *Journal of Clinical Psychopharmacology, 34*(3), 297–302. doi:10.1097/JCP.0000000000000112

Tedeschini, E., Fava, M., & Papakostas, G. I. (2011). Placebo-controlled, antidepressant clinical trials cannot be shortened to less than 4 weeks' duration: A pooled analysis of randomized clinical trials employing a diagnostic odds ratio-based approach. *The Journal of Clinical Psychiatry, 72*(1), 98–113. doi:10.4088/JCP.09m05725blu

Tohen, M., Bowden, C. L., Nierenberg, A. A., & Geddes, J. (2015). *Clinical trial design challenges in mood disorders*. Amsterdam, The Netherlands/Boston, MA: Academic Press.

Trevino, K., McClintock, S. M., McDonald Fischer, N., Vora, A., & Husain, M. M. (2014). Defining treatment-resistant depression: A comprehensive review of the literature. *Annals of Clinical Psychiatry, 26*(3), 222–232.

Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T., Crismon, M. L., Shores-Wilson, K., Toprac, M. G., Dennehy, E. B., Witte, B., & Kashner, T. M. (2004). The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: A psychometric evaluation. *Psychological Medicine, 34*(1), 73–82. doi:10.1017/S0033291703001107

Uher, R., Farmer, A., Maier, W., Rietschel, M., Hauser, J., Marusic, A., Mors, O., Elkin, A., Williamson, R. J., Schmael, C., Henigsberg, N., Perez, J., Mendlewicz, J., Janzing, J. G., Zobel, A., Skibinska, M., Kozel, D., Stamp, A. S., Bajs, M., Placentino, A., Barreto, M., McGuffin, P., & Aitchison, K. J. (2008). Measuring depression: Comparison and integration of three scales in the GENDEP study. *Psychological Medicine, 38*(2), 289–300. doi:10.1017/S0033291707001730

Uher, R., Perlis, R. H., Placentino, A., Dernovšek, M. Z., Henigsberg, N., Mors, O., Maier, W., McGuffin, P., & Farmer, A. (2012). Self-report and clinician-rated measures of depression severity: Can one replace the other? *Depression and Anxiety, 29*(12), 1043–1049. doi:10.1002/da.21993

Wilcox, C. S., Cohn, J. B., Linden, R. D., Heiser, J. F., Lucas, P. B., Morgan, D. L., & DeFrancisco, D. (1992). Predictors of placebo response: A retrospective analysis. *Psychopharmacology Bulletin, 28*(2), 157–162.

Yamauchi, M., Miyara, T., Matsushima, T., & Imanishi, T. (2006). Desensitization of 5-HT2A receptor function by chronic administration of selective serotonin reuptake inhibitors. *Brain Research, 1067*(1), 164–169. doi:10.1016/j.brainres.2005.10.075

Zanarini, M. C., Skodol, A. E., Bender, D., Dolan, R., Sanislow, C., Schaefer, E., Morey, L. C., Grilo, C. M., Shea, M. T., McGlashan, T. H., & Gunderson, J. G. (2000). The collaborative longitudinal personality disorders study: Reliability of axis I and II diagnoses. *Journal of Personality Disorders, 14*(4), 291–299. doi:10.1521/pedi.2000.14.4.291

Zarate, C. A., Singh, J. B., Carlson, P. J., Brutsche, N. E., Ameli, R., Luckenbaugh, D. A., Charney, D. S., & Manji, H. K. (2006). A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. *Archives of General Psychiatry, 63*(8), 856–864. doi:10.1001/archpsyc.63.8.856

Zimmerman, M., Walsh, E., Friedman, M., Boerescu, D. A., & Attiullah, N. (2017). Identifying remission from depression on 3 self-report scales. *The Journal of Clinical Psychiatry, 78*(2), 177–183. doi:10.4088/JCP.16m10641