



# Great Expectations: recommendations for improving the methodological rigor of psychedelic clinical trials

Jacob S. Aday<sup>1</sup> · Boris D. Heifets<sup>2</sup> · Steven D. Pratscher<sup>3</sup> · Ellen Bradley<sup>1</sup> · Raymond Rosen<sup>1</sup> · Joshua D. Woolley<sup>1</sup>

Received: 18 October 2021 / Accepted: 14 March 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

**Rationale** Psychedelic research continues to garner significant public and scientific interest with a growing number of clinical studies examining a wide range of conditions and disorders. However, expectancy effects and effective condition masking have been raised as critical limitations to the interpretability of the research.

**Objective** In this article, we review the many methodological challenges of conducting psychedelic clinical trials and provide recommendations for improving the rigor of future research.

**Results** Although some challenges are shared with psychotherapy and pharmacology trials more broadly, psychedelic clinical trials have to contend with several unique sources of potential bias. The subjective effects of a high-dose psychedelic are often so pronounced that it is difficult to mask participants to their treatment condition; the significant hype from positive media coverage on the clinical potential of psychedelics influences participants' expectations for treatment benefit; and participant unmasking and treatment expectations can interact in such a way that makes psychedelic therapy highly susceptible to large placebo and nocebo effects. Specific recommendations to increase the success of masking procedures and reduce the influence of participant expectancies are discussed in the context of study development, participant recruitment and selection, incomplete disclosure of the study design, choice of active placebo condition, as well as the measurement of participant expectations and masking efficacy.

**Conclusion** Incorporating the recommended design elements is intended to reduce the risk of bias in psychedelic clinical trials and thereby increases the ability to discern treatment-specific effects of psychedelic therapy.

**Keywords** Psychedelics · Psychedelic therapy · Clinical trials · Treatment expectations · Expectancies · Placebo effect · Masking · Recommendations

## Overview

Recent high-profile clinical trials with psychedelic drugs have highlighted challenges related to rigorous study design and condition masking that have simmered in both psychotherapy and pharmacology research for decades (e.g.,

---

Boris D. Heifets is co-first author

---

This article belongs to a Special Issue on Psychopharmacology on Psychedelic Drugs

---

✉ Jacob S. Aday  
jacob.aday@ucsf.edu

Boris D. Heifets  
bheifets@stanford.edu

Steven D. Pratscher  
spratscher@ufl.edu

Ellen Bradley  
Ellen.Bradley@ucsf.edu

Raymond Rosen  
raymondrosen@gmail.com

Joshua D. Woolley  
josh.woolley@ucsf.edu

<sup>1</sup> Department of Psychiatry and Behavioral Sciences, San Francisco VA Medical Center, University of California, 401 Parnassus Ave., San Francisco, CA 94143, USA

<sup>2</sup> Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University, Stanford, CA, USA

<sup>3</sup> Department of Community Dentistry and Behavioral Science, Pain Research and Intervention Center of Excellence, University of Florida, Gainesville, FL, USA

**Table 1** Key terms and definitions

Key terms	Definition
Confounding variable	A factor other than variables under study that influences the dependent variable
Hawthorne effect	Changing one's behavior as response to the interest, care, or attention received through observation and assessment
Regression to the mean	Tendency for extreme scores to return to average over time
Spontaneous remission	Symptom resolution independent of the study manipulation as a function of an unidentified biological or psychosocial change
Process expectancies	Expectations regarding what will happen during the treatment
Outcome expectancies	Expectations regarding the outcome of the treatment
Active placebo	A control condition that closely resembles the presentation and side effects of the experimental treatment without providing the therapeutic effects
Masking efficacy	Degree to which participants are unaware of their treatment arm assignment
Placebo effect	Treatment-nonspecific improvement in symptoms attributable to contextual factors, such as participants' positive expectations regarding the treatment
Nocebo effect	Treatment-nonspecific worsening of symptoms as a result of contextual factors, such as negative expectations regarding the treatment
Michael Pollan Effect	Heightened positive expectations regarding the efficacy of psychedelics in recent years stemming from Michael Pollan's book, <i>How to Change Your Mind</i>

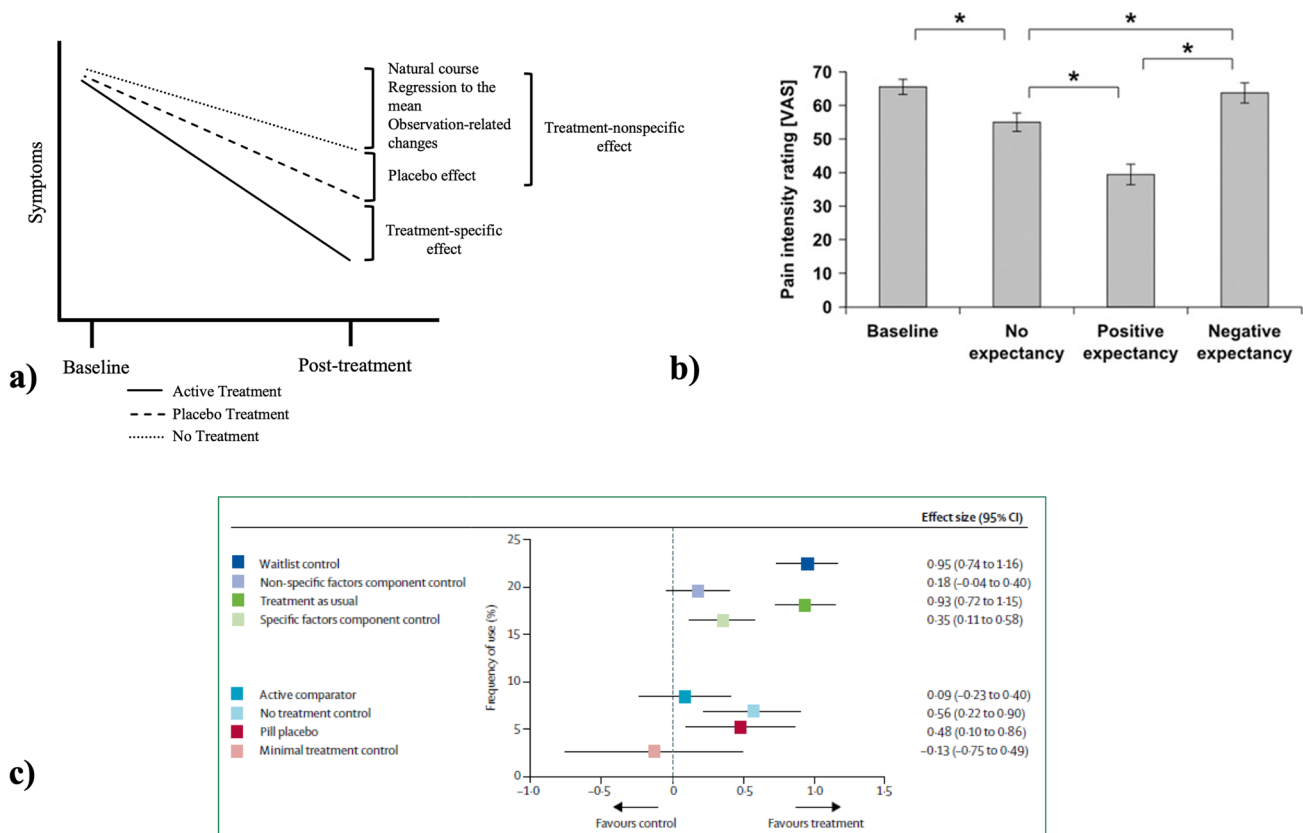
Basoglu et al. 1997; Enck and Zipfel 2019). Interrelated methodological challenges regarding the selection of appropriate control conditions, masking (also known as blinding<sup>1</sup>), and expectancy effects have clouded our understanding of the source of clinical improvements in psychedelic studies and, in fact, across medicine. Studies on psychedelic therapy are particularly challenging as they must address methodological issues inherent to both psychotherapy and pharmacology research as well as issues that are distinctly problematic to the field, such as “hype” and salient psychoactive effects that compromise masking. In this paper, we delineate how many of the methodological limitations that have been raised as critiques of psychedelic science are common challenges across psychotherapy and pharmacology research more broadly and are in need of addressing. This review allows us to share lessons across disciplines and provide recommendations for improving future psychedelic and non-psychedelic research. We conclude by highlighting that psychedelic studies should not be held to a different standard than other forms of psychotherapy or pharmacology research, and that the fields can leverage important lessons from one another by recognizing their shared limitations. To this end, we provide practical methodological recommendations to measure and manage expectations as well as to enhance masking in psychedelic studies. These recommendations can be deployed more broadly across clinical trials to improve the rigor and reproducibility of future research.

## Treatment-nonspecific effects

To begin, we review the various reasons for including control conditions in clinical studies and examine what exactly is being controlled. In any clinical trial, changes in symptoms can be observed because of treatment-specific or treatment-nonspecific effects (Turner et al. 1994). Treatment-specific effects are changes directly attributable to the independent variable or intervention under study (e.g., drug dose or psychotherapeutic approach). Treatment-nonspecific effects are changes not related to the specific treatment arm (i.e., common to being in any clinical trial), as well as placebo and nocebo effects related to treatment expectations (Table 1). Including certain control conditions allows the trialist to filter out contributions of treatment-nonspecific effects from treatment-specific effects to attribute clinical improvements to the intervention under study (Fig. 1a).

The natural history or spontaneous variation of any given disease under study may be the least controllable source of treatment-nonspecific change that can confound clinical trial interpretation. Symptoms can change (e.g., spontaneous remission) independently of the study intervention as a function of an unidentified biological or psychosocial change in the individual's life. Additionally, in most clinical trials, participants are screened and selected based on minimum criteria of symptom severity, and many individuals may be especially motivated to seek out research studies when their symptoms peak in severity (Whitney and Von Korff 1992). Subsequent measurements using the same scale may show an apparent improvement. This “regression to the mean” rather than a true treatment-specific effect may lead researchers to erroneously conclude a treatment is effective when participants may have improved over time without

<sup>1</sup> In recent years, the term “masking” has been used in place of “blinding”; here, we have opted to use the term “masking” but consider the terms synonymous



**Fig. 1** Treatment-nonspecific effects in clinical trials. **(a)** Hypothetical results of a clinical trial to delineate the sources of treatment-specific and treatment-nonspecific effects. Including placebo and no treatment control conditions allows trialists to identify treatment-specific effects (figure inspired by Wampold et al. 2016). **(b)** In a clear illustration of expectancy effects, Bingel et al. (2011) measured participants’ pain intensities before (i.e., Baseline) and after receiving remifentanyl while manipulating participant expectancies across

three groups (e.g., No expectancy, Positive expectancy, or Negative expectancy). They found that priming positive treatment expectancy doubled the analgesic effect of remifentanyl when compared to no expectancy. In contrast, inducing negative treatment expectancies eliminated the analgesic effect. **(c)** Gold et al. (2017) demonstrated that treatment effect sizes vary as a function of the type control group utilized

any treatment (Hengartner 2020). Regression to the mean is a ubiquitous statistical phenomenon that results whenever cases are selected for follow-up based on abnormally high or low scores at baseline, demonstrated in observational studies and clinical trials, and across multiple diseases (Bland and Altman 1994). Changes due to the natural course of the condition and regression to the mean are considered theoretically distinct but in practice are difficult to disentangle.

Participant behavior can also change simply as a consequence of the interest, care, or attention received as part of a study. This well-established psychological phenomenon is known as the Hawthorne effect (Sedgwick and Greenwood 2015). This effect is associated with outcomes as diverse as workplace productivity to cognitive functioning and quality of life in dementia patients (McCarney et al. 2007). Notably, researchers and study personnel, not just participants, can be susceptible to Hawthorne effects, thereby influencing clinical outcomes (Sedgwick and Greenwood 2015). That is, those caring for participants in an experimental trial are

under increased scrutiny and observation as compared to those operating in an unobserved clinical setting, and this difference may impact both the quality and quantity of patient care. This bias can cause an overestimation of an experimental treatment’s therapeutic effect due to clinical improvements from treatment-nonspecific factors. A distinct but related issue is that the simple act of repeated observation and measurement of behaviors and symptoms can alter those same behaviors and symptoms. Repeated pain assessments can increase pain chronicity (Ferrari and Russell 2010), asking about illicit drug use can decrease use (D’Onofrio et al. 2012), and daily symptom assessments can worsen or improve symptom severity in PTSD (Dewey et al. 2015; Pedersen et al. 2014). Drawing extra attention to an issue can lead to symptom amplification or may provide more opportunities to resolve it (Barsky and Klerman 1983). In either case, it is clear that simply enrolling in a clinical trial can influence symptoms regardless of treatment assignment.

Taken together, issues related to the natural course of the disease, regression to the mean, and observation-related changes highlight that there are many mechanisms by which symptoms may change in a clinical trial irrespective of the treatment being tested. It is therefore important to include, at a minimum, control arms that do not receive the treatment, as treatment-nonspecific factors confound experimental and control arms to a similar extent. However, the simple inclusion of an untreated comparison group may not be enough to isolate treatment-specific effects (Gold et al. 2017; Enck and Zipfel 2019). Participants often have expectations regarding the efficacy of the treatment under study. If participants have knowledge about their treatment arm assignment (e.g., in an open-label study), or gain knowledge through their subjective experience (e.g., having a psychedelic trip) or somatic symptoms, their expectations about therapeutic efficacy can affect their clinical outcomes. This problem is common to most psychotropic trials (e.g., selective serotonin reuptake inhibitors [SSRIs]; Hieronymus et al. 2018) and is particularly salient for high-dose psychedelic trials in which subjective drug effects are especially pronounced. Without effective condition masking, it is virtually impossible to maintain the independence of the main variable under study (i.e., the treatment), as it is confounded by participant expectations. In addition to influencing participant outcomes, baseline expectancies about a treatment's therapeutic effects can also impact masking efficacy (i.e., whether participants are aware of their treatment arm assignment), as those with noticeable improvements in symptoms often assume they were assigned to the active treatment group (Sackett 2007). We now consider several specific expectations and how they interact with masking and treatment outcomes.

## Expectancies in psychotherapy and pharmacology research

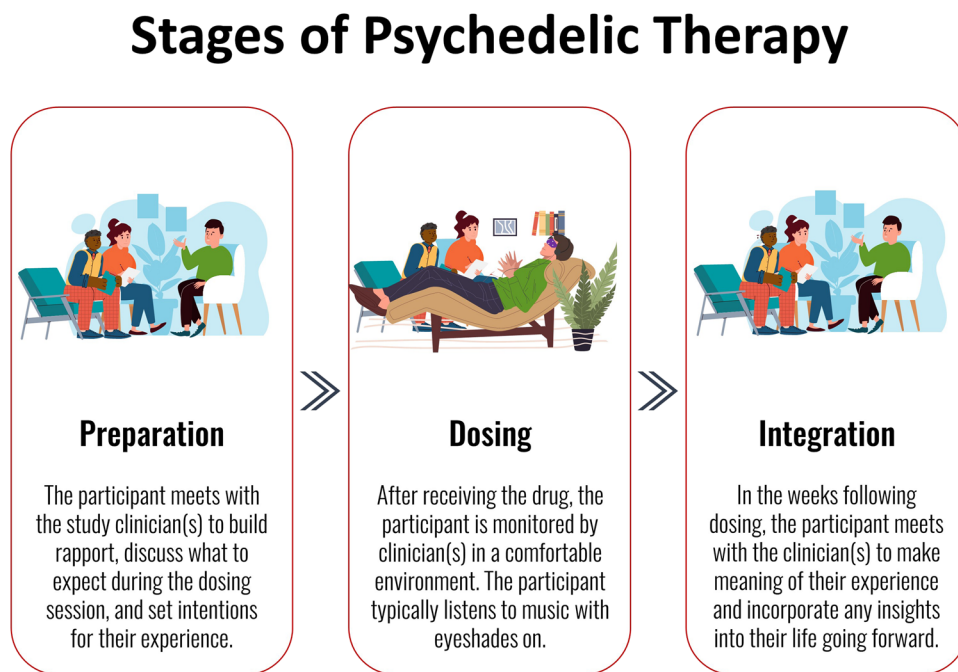
Tambling (2012) differentiates between expectations about the process of treatment and expectations about the outcome of treatment. In the case of psychotherapy, process expectations are expectations about what will happen during therapy (e.g., patient's thoughts about roles they and their therapist will assume, characteristics of their therapist, and what sessions will entail). In pharmacological trials, process expectations can include expectations about any acute drug effects, including psychoactive effects. Process expectancies may be particularly pertinent with psychedelic drug trials as expectations about the acute effects of the drugs are shaped by hours of psychotherapy, widespread representations in popular media, and a highly ritualized process of drug administration. When

these expectations are matched by experience, a study participant may be especially confident in unmasking their treatment arm assignment.

Outcome expectations refer to whether the treatment is anticipated to reduce symptoms. In the case of psychotherapy, studies suggest that outcome expectancies are stronger predictors of therapeutic effects than are specific psychotherapy techniques (Horvath et al. 2011; Webb et al. 2010). Positive outcome expectations are related to stronger alliance with the therapist, which is associated with better outcomes (Višlă et al. 2018; Yoo et al. 2014). A recent, well-powered meta-analysis ( $N = 12,722$ ) compared patient outcome expectancies and clinical outcomes across a variety of diagnoses and psychotherapy interventions, revealing that greater positive outcome expectancy was consistently associated with better treatment results (Cohen's  $d = 0.36$ ; Constantino et al. 2018). Outcome expectancies also have strong effects relative to the active effects of psychotropic drugs (Rutherford and Roose 2013). In trials where patient-reported outcomes are the primary efficacy measures, the effects of outcome expectancies are particularly strong (Atlas 2021). Fillingim and Price (2005) concluded that in placebo analgesia studies outcome expectancies accounted for up to 81% of variance in post-treatment pain ratings. Thus, across clinical research contexts, participants' outcome expectations about the specific treatment being administered influence clinical outcomes.

Negative outcome expectations can also influence clinical outcomes. When individuals are aware that they have been assigned to a treatment that they believe is unlikely to improve their symptoms, negative expectation alone can worsen patient outcomes, which is known as the nocebo effect (Gold et al. 2017; Planès et al. 2016). This effect was elegantly demonstrated in a study with remifentanyl, an opioid analgesic, which found that priming negative expectations about the treatment completely negated the analgesic effect of the drug (Bingel et al. 2011; Fig. 1b). Furthermore, if a participant has positive expectations about the proposed experimental treatment but comes to believe they have been assigned to a control condition, outcomes may worsen as a result of disappointment or the belief that one will not improve without being assigned to the active treatment (Furukawa et al. 2014). Indeed, those put on a waitlist control condition typically have worse outcomes than those assigned to active placebo, or even no treatment, as they have less reason to expect an improvement in symptoms (Patterson et al. 2016). With waitlist control designs, those in the control condition do not receive treatment until after a waiting period, where they are compared with the active treatment group. However, participants are generally aware that they are in a control condition during their waiting period and thus may not expect to see improvements, whereas the

**Fig. 2** Stages of psychedelic therapy. Psychedelic therapy typically involves preparation, dosing, and integration sessions



active treatment group likely has the opposite expectation. Therefore, waitlist control designs may artificially inflate intervention effect size estimates (Fig. 1c; Cunningham et al. 2013; Zhipei et al. 2014). Possibly illustrating this effect, in a waitlist control study of psilocybin for the treatment of major depressive disorder, waitlisted participants reported higher anxiety scores at the end of the waitlist period compared to the beginning, enhancing the apparent therapeutic effect of psilocybin (Davis et al. 2021). The crucial role of expectancies in treatment outcomes across clinical contexts underscores the need for trial designs that control for expectation-related improvements, which we elaborate on in the following sections.

Importantly, outcome expectancies are rarely measured in psychotherapy and pharmacology studies (Doering et al. 2014). Constantino et al. (2011) noted that expectancies have often been thought of as nuisances to clinical research and disregarded rather than being considered important ingredients of the therapeutic process. Furthermore, the few studies that have included assessments of treatment expectations have used brief and study-specific measures, meaning there is surprisingly little overlap between studies in how expectations are quantified (Tambling 2012). Moreover, there is no manual or expert consensus for managing expectancies despite the extensive evidence of the important role of expectancy in treatment responses (Zilcha-Mano et al. 2019). Collectively, these findings highlight that challenges related to participant expectations are common across psychotherapy and pharmacology research, and that, to date, there is no standard for addressing expectation-related issues.

### Psychedelic research and expectations

Briefly, the typical structure of a modern psychedelic therapy clinical trial involves an arduous screening process, multiple preparation sessions, single or multiple drug dosing sessions, and integration sessions after drug administration (Fig. 2). The preparation sessions are used for several purposes, including to build rapport between the participant and the therapists or facilitators<sup>2</sup>, to inform the participant about common or possible psychedelic drug experiences, to reassure the participant's safety with dosing day procedures, and to assist with establishing the patient's intention(s) for their dosing session. The drug dosing session is highly structured with two therapists accompanying the participant throughout the 6–8-h session in a comfortable environment. During the dosing session, participants often remain reclined on a couch with eyeshades and headphones for music and are encouraged to focus on their inner experience throughout the drug session, exploring any content that arises with an open and accepting mindset. In the days following drug dosing, the participants work with the same clinical team in integration

<sup>2</sup> Notably, there is significant debate about the proper terminology for the people who provide the preparation and integration and who monitor participants during the dosing session. “Guide,” “sitter,” “facilitator,” “therapist,” “monitor,” and other terms have been proposed and have their advocates and detractors. The intensity of these debates highlights the truth of the old joke that “Scientists would rather use each other's toothbrushes than use each other's terminology.” We use the term “facilitator” throughout this manuscript without taking a strong stance on which term is the most correct

sessions to make meaning of their experiences and to incorporate any insights they may have had into their lives going forward. With these fundamental elements of psychedelic therapy, it is best considered a complex, multicomponent intervention that includes aspects of both pharmacology and psychotherapy. Notably, throughout the course of a psychedelic therapy trial, a participant's process expectations and outcome expectations are subject to change as they gather more information about possible drug effects, approach the sessions in a certain way (e.g., trust, let go, be open), and experience the actual drug effects. Hereafter, we refer to this package of procedures as psychedelic therapy and acknowledge that all of these aspects may determine treatment-specific effects.

Participants' expectations as well as intentions (i.e., what they desire from the psychedelic experience) are thought to play a prominent role in the drugs' acute and long-term effects (Olson et al. 2020). Some have even termed psychedelics "placebo enhancers," as they can enhance the perception of meaningfulness (Hartogsohn 2016, 2018) and induce a state of suggestibility (Carhart-Harris et al. 2015). It has been noted across popular culture that psychedelic experiences are heavily influenced by one's expectations, and some have gone as far as to claim "no other class of drugs are more suggestible in their effects" (Pollan 2018). Hartogsohn (2021) noted that the fundamental role of expectations in psychedelic drug effects may reconcile the paradoxical conceptions that have been held about the drugs—views that are so varied, it at times sounds as though scientists are discussing completely different drugs (e.g., they have been used to both treat mental illness and to model psychosis). Utilizing pre-dosing expectations as well as the acute state of suggestibility induced by psychedelics in tandem may be an important component of the therapeutic process with psychedelic therapy, but this combination can also be co-opted for nefarious purposes. Historically, psychedelics have been used by cults as well as investigated for their alleged potential in "mind control" by the US government during MK Ultra (Cusack 2020; Kogo 2002; Ledford 2019). There is even concern about psychedelics' potential for changing beliefs (e.g., political or metaphysical; de Wit et al. 2021; Pace and Devenot 2021; Timmermann et al. 2021) and memories, though that is beyond the scope of this review. Therefore, it may be ethical to include an enhanced informed consent process about possible belief changes induced by psychedelic therapy prior to enrolling participants into a clinical trial (Smith and Sisti 2021).

Although pre-dosing expectations have long been thought to be integral to the effects of psychedelics (Eisner 1997; Leary et al. 1963), very few studies have actually measured them. A recent "microdosing" (i.e., sub-hallucinogenic dosing) study found that positive expectations regarding psychedelics at baseline predicted subsequent increases in

wellbeing irrespective of whether a participant received a psychedelic or an inert placebo (Kaertner et al. 2021). Similarly, a large-scale, placebo-controlled study of microdosing found that participants experienced comparable improvements in mood and cognition in the drug and placebo conditions (Szigeti et al. 2021). Another microdosing study found that after controlling for baseline expectancies, there was no difference between psilocybin and placebo on measures of awe (van Elk et al. 2021). However, to the best of our knowledge, only a single "macro-dosing" (i.e., full hallucinogenic dosing) trial has recorded pre-treatment expectancies. An open-label ayahuasca study found that participants endorsing an expectancy of favorable change in neuroticism, extraversion, and conscientiousness in response to ayahuasca showed a greater decrease in neuroticism and greater increases in extraversion and conscientiousness following ayahuasca administration compared to participants with lower expectancies receiving the same treatment (Weiss et al. 2021). A recent systematic review found those with a recreational intention with psychedelics tended to have less challenging experiences when they used a psychedelic (Aday et al. 2021; Haijen et al. 2018), again suggesting that what one desires and expects to experience with psychedelic influences the drug's effects. Thus, the few studies that have measured expectations and intentions to date support the prevalent assumption that pre-dosing expectations interact with psychedelic drug effects and outcomes. Whether these same considerations apply to other drug classes (e.g., such as psychostimulants) is unknown, further emphasizing the need to measure and report therapeutic expectations in a systematic way across areas of clinical research.

High-dose psychedelic trials may also be particularly susceptible to a type of bias termed "hype" or the "Michael Pollan effect" (Carpenter 2020; Table 1). Some have argued that psychedelic therapy marks the most important innovation in psychiatry since the introduction of SSRIs, or possibly ever, and it is not uncommon to hear claims about the potential for psychedelics to "change the world" from industry leaders and enthusiasts (Dupuis 2021). This pervasive messaging may lead to amplified positive expectations compared with many other types of clinical interventions and perhaps motivates participants to "not let the movement down" by failing to clinically improve. This notion was illustrated in a recent ayahuasca study (Aday 2021), where one of the participants asked us (JSA) if they should stop participating in the study because they did not have a mystical experience and did not want to "ruin the research." In our experience recruiting for psychedelic studies, many potential participants explicitly express a sense of pride and excitement in participating in a psychedelic trial as well as strong confidence in the benefit of psychedelics to their mental wellbeing. These motivations for participation and heightened positive expectations coupled with the functional unmasking that often occurs make

identification of a treatment-specific effect in high-dose psychedelic trials particularly challenging and highlights the need for study designs that properly mask participants to conditions (Burke and Blumberger 2021).

Certain aspects of the study personnel, environmental context, and measures included in psychedelic drug trials may contribute to enhanced expectations as well. For example, the use of two therapists at a time and rituals like placing a fresh rose in the room on dosing day may serve to amplify positive expectations and signal that the experience is of particular significance (Gukasyan and Nayak 2021). Additionally, outcome expectancies of psychotherapists have been shown to have a marked effect on treatment engagement and clinical outcomes across therapeutic approaches (Doering et al. 2014; Leake and King 1977), suggesting this may be a treatment-nonspecific factor relevant to psychedelic studies as well. Lastly, the specific measures used in psychedelic trials can influence participant expectations; one study volunteer noted “I long to see some of the stuff hinted at in the questionnaire” in reference to questions they encountered on the Mystical Experience Questionnaire (MEQ; MacLean et al. 2012; Pollan 2018). Thus, in addition to preexisting attitudes about psychedelics, certain expectations may be engendered by characteristics of the trial.

## Modern era clinical research design elements

Next, we will describe many of the study designs and methods that have been attempted to manage these issues across psychotherapy and pharmacology trials to date. Open-label study designs, in which both the patient and study personnel are aware of what specific treatment is administered, most closely resemble how psychotherapy and psychotropic drugs are administered in real-world, non-research settings. Although high in ecological validity, this type of design does not control for most of the confounding nonspecific factors that can affect clinical outcomes (e.g., Hawthorne effect, spontaneous variation of symptoms, regression to the mean).

Some treatment-nonspecific factors, such as regression to the mean, can be controlled if sufficient data are available at both the individual and group level, as a precise mathematical formula can be developed to predict the actual regression effect in a given experimental setting (Barnett et al. 2005). These authors have identified specific experimental strategies to mitigate or manage expected regression to the mean effects in a clinical trial. First, they recommend selecting cases based on multiple baseline observations. Requiring that eligible subjects have stable test scores over two or more baseline assessments will predictably reduce, although not necessarily eliminate, regression to the mean. Second, the authors suggest correcting for regression to the mean

effects in the analyses by using either ANCOVA modeling or application of a correction formula. Of note, neither of these strategies have been systematically applied in studies of psychedelic therapy. Third, investigators may consider a waitlist control condition, although we refer the reader to limitations to this approach noted previously.

The double-blind randomized controlled trial (RCT) is considered the gold standard design for identifying a true treatment-specific effect, under conditions where neither investigator nor participant knows their treatment allocation. An RCT entails randomly assigning participants to treatment or control conditions and withholding knowledge of treatment arm assignment from participants and study personnel (i.e., masking). Effectively executing this design controls for expectancies as it is unknown which treatment each participant received, and therefore treatment-nonspecific factors can be ruled out as the source of treatment arm outcome differences. Treatment arm masking in RCTs is best achieved with active placebo comparators, in which the control condition is structurally equivalent and closely resembles the presentation and side effects of the experimental treatment without providing the therapeutic effects (Doering et al. 2014). Inert but identical-looking pills that lack the side effects of the treatment condition (i.e., inactive placebos) are often used but may be easy for participants to detect, and subsequent nocebo effects may confound analyses.

There has been considerable debate that continues today about what constitutes a proper “inert” placebo for psychotherapy in the same sense as an “inert” placebo in pharmacology, as some have argued that “there is no such thing as inert psychotherapy” (Rosenthal and Frank 1956; Wampold et al. 2016). In the context of psychedelic trials, to date, the psychotherapy component has been held constant across the treatment and control conditions, making this issue less relevant for the field for now. However, as researchers delineate the nuances of what specific forms of psychotherapy are most synergistic with psychedelics, this potential confound will become an increasingly important issue to address (Horton et al. 2021). A related challenge with psychedelic studies is that unmasking may lead to differences in how the psychotherapy component is administered and received, given that the context of the therapy shifts once the participant and/or therapist becomes aware of the treatment arm assignment. Therefore, improved masking procedures must be implemented into psychedelic science for the field to meet the assumptions of the current gold standard clinical trial design.

Crossover RCT designs have been used in many pharmacological studies as an efficient way to account for treatment-nonspecific confounds because participants act as their own control. In a crossover design, participants are randomly assigned to a sequence of treatments where they receive both the experimental and placebo treatments but at different

timepoints (i.e., placebo then experimental treatment or vice versa). A major weakness of crossover designs, however, is the potential for carryover effects (i.e., the therapeutic benefits could “carryover” after the first treatment and misrepresent the true effect of the second treatment). Carryover effects are especially concerning in psychedelic trials because the effects of psychedelic therapy in some cases have been shown to be durable for over a year (Griffiths et al. 2008; Johnson et al. 2017; see Aday et al. 2020b for review). Thus, even a 12-month washout period is unlikely to achieve a return to pre-treatment levels on the variable of interest, which biases within-person analyses and threatens the validity of conclusions that can be drawn. Moreover, masking is likely to be compromised in crossover designs that involve a psychoactive drug (Wilsey et al. 2016). For example, almost all participants accurately identified their treatment condition in a crossover study that used psilocybin and niacin as a placebo control (Grob et al. 2011). Thus, simple crossover designs may be more confounded than a parallel (between-subjects) RCT design for psychedelic trials.

We have repeatedly noted the importance of adequate masking in double-blind RCTs, and emphasize that it is impossible to know if the double-blind or masking was achieved without testing masking efficacy. Surprisingly, however, masking efficacy typically goes unmeasured or unreported in psychotherapy and pharmacology trials (Doering et al. 2014). Many researchers report their studies as being “double-blind” without testing such claims (Basoglu et al. 1997). A systematic review on methods of masking in randomized controlled trials with pharmacologic treatments concluded that reporting of condition masking is generally “quite poor,” and based on trials that have tested the success of masked methods, a high proportion of studies are effectively unmasked (Boutron et al. 2006; Rabkin et al. 1986). This corroborates a recent systematic review of studies published in top psychiatry journals in 2017 and 2018, which found that only 59% of the trial reports included adequate reporting of masking outcomes (Juul et al. 2020), as well as a meta-analysis that indicated a large majority of antidepressant RCTs do not assess masking efficacy, and when measured, masking often fails (Scott et al. 2022). Similarly, a comprehensive literature search found that masking was not maintained in 20/23 “double-blind” studies examining psychotropic drugs (Fisher and Greenberg 1993). The authors noted improvements in patient symptomatology and side effects from the active drug were the major cause of unmasking. Long-term masking can be difficult, if not impossible, to achieve with highly efficacious treatments because it is clear to the patient that they experienced an improvement in symptoms (Muthukumaraswamy et al. 2021). Thus, many argue that end-of-trial assessments for masking cannot be done with validity, as they cannot disentangle masking from guesses based on efficacy (Mataix-Cols

and Andersson 2021; Sackett 2007), although it should be noted that some researchers argue that it is not considered unmasking at the end of the trial if people guess their condition based on efficacy (Katz 2021).

## Masking attempts in psychedelic studies

Multiple approaches have been attempted to address these methodological challenges specifically as they relate to psychedelic trials. First, active placebos have been used in an attempt to mask participants and therapists to treatment conditions, albeit generally unsuccessfully. This difficulty was infamously demonstrated in the “Good Friday Experiment,” where divinity school students were assigned to receive psilocybin or niacin, a B vitamin with mild physiological effects, in a group setting at a chapel (Pahnke 1963). Despite some initial confusion because of niacin’s fast-acting effects on vasodilation and general relaxation, before long, it became clear which participants had been assigned to which condition, as those in the psilocybin group had intense subjective reactions and often spiritual experiences, whereas the niacin group “twiddled their thumbs” while watching on (Prideaux 2021). By the end of the day, all participants correctly ascertained whether they were in the treatment or control group (Doblin 1991). Despite the clear masking failure, after more than 50 years, many researchers today still use niacin as the active placebo in clinical trials with psychedelics, perhaps for a lack of better alternatives (Grob et al. 2011; Ross et al. 2016; Siegel et al. 2021). Nevertheless, participants are now dosed individually rather than in a group to reduce potential unmasking from witnessing others’ experiences. Modern psilocybin trials have also employed methylphenidate (Griffiths et al. 2006) and dextromethorphan (DXM; Carbonaro et al. 2018) as active placebos, although the success of masking was typically less than 25% or unreported in these studies (Bershad et al. 2019; Carbonaro et al. 2018; Griffiths et al. 2006). Uthaug et al. (2021) tested an innovative strategy at masking by mimicking the aesthetic and somatic features of the psychedelic brew, ayahuasca. The investigators used a mixture of coco powder, vitamins (unspecified), turmeric powder, quinoa, traces of coffee, and potato flour, as a placebo to mimic the texture as well as gastrointestinal side effects of the drug. Despite effectively masking the profound effects of ayahuasca in several experienced users, a majority of participants were still able to accurately identify their treatment assignment (Uthaug et al. 2021). A review of ongoing clinical trials revealed that researchers are currently experimenting with a number of other potential control conditions in psychedelic studies, including mannitol, lactose, ketamine, microcrystalline cellulose, and nicotinamide (Siegel et al. 2021), but the effectiveness of these attempts remains to be seen.



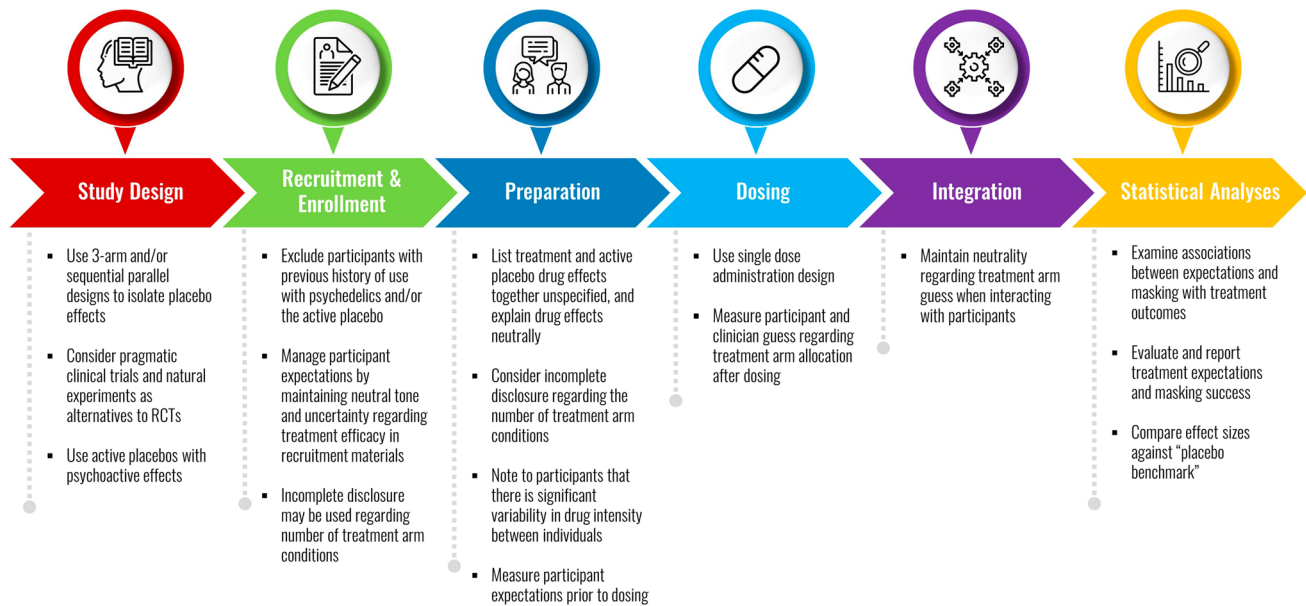
Low doses of psychedelics have also been tried as a potential control condition to improve participant masking (Griffiths et al. 2016). One study combined a low dose of psilocybin with incomplete disclosure (see below) such that participants and study staff were unaware of the number of treatment arms in the study. Specifically, participants were informed that they could receive anywhere from 0.5 to 30 mg of psilocybin in the trial when in fact they could only receive 0.5 mg if they were in the control condition or 25 mg if they were in the treatment condition (Griffiths et al. 2016). An advantage of including the low dose of psilocybin is that all participants are truthfully told they will receive psilocybin, which presumably helps balance treatment expectations across both conditions. However, participants and therapists are still at risk for unmasking with this design because it is typically easy to ascertain whether the participant has an intense psychedelic experience or not. Schenberg (2021) also noted that this design may be limited by ethical considerations, given that 3,4-methylenedioxymethamphetamine (MDMA) research has shown that low-dose control conditions can be stressful and trying for patients, leading to drop-outs and dissatisfaction (Oehen et al. 2013), and anecdotal lore in the underground psychedelic therapy community suggests that medium doses of psychedelics can agitate people without allowing them to “breakthrough” (JDW, personal communication, 2021). On the other hand, low doses of classic psychedelics (i.e., microdosing) have been purported to be therapeutic (Fadiman 2011; Kuypers et al. 2019), which could also confound study results, although the therapeutic benefit of single microdoses seems unlikely to be durable or significant. Thus, including a low-dose psychedelic as part of an active control condition is a promising starting point.

Incomplete disclosure of certain aspects of the study design is a strategy that has been employed to enhance masking success and balance treatment expectations among conditions. For example, some studies incompletely disclose the number of treatment arms to participants in an attempt to obscure the study design and reduce the participants’ confidence in their treatment group allocation (Bershad et al. 2019; Carbonaro et al. 2018; Griffiths et al. 2006; Reissig et al. 2012). Another compelling approach (in healthy subjects) involves consenting participants to possibly receiving one of several substances in order to reduce their certainty of treatment allocation. For example, in some experiments, participants consent to receive MDMA, methamphetamine, tetrahydrocannabinol (THC), benzodiazepine, and/or placebo (Bedi et al. 2010; Bershad et al. 2019), but in fact only receive one or two of these drugs in any particular study. Although this design is possible to implement in psychedelic studies of healthy individuals who are not seeking treatment, there are limitations to this approach, including reduced generalizability because a large proportion of the population may not be comfortable with receiving any

one of the listed substances. Moreover, this design has not proven to be particularly effective to date, as participants accurately identify the experimental condition (e.g., MDMA and psilocybin) ~70–85% of the time (Bershad et al. 2019; Carbonaro et al. 2018). Thus, even with these more rigorous approaches, adequate masking remains a challenge. Taken together, there is a pressing need for methodological innovations that adequately address the problem of masking in psychedelic studies.

Muthukumaraswamy et al. (2021) made several recommendations for addressing masking in psychedelic clinical trials. The authors suggested that active placebos may need to be combined with alternative trial designs (e.g., dose-response parallel-groups design) as well as some vagueness about the acute effects of psychedelics when consenting participants. Dose-response parallel-groups designs compare the full dose of the active treatment drug with a low dose; the advantages and disadvantages of such an approach are discussed previously. Vagueness regarding the acute effects of psychedelics has tradeoffs as well: although it may improve masking, there are clear ethical concerns as participants need to be able to give fully informed consent (Smith and Sisti 2021). This consideration is especially true with psychedelic studies, as psychedelic experiences have been described as “life changing” and have the potential to affect one’s social relationships (Ross et al. 2016), spirituality (Griffiths et al. 2006), and worldview (Timmermann et al. 2021). Another recommendation provided was the  $2 \times 2$  balanced placebo design (Rohsenow and Marlatt 1981), or  $2 \times 2$  factorial design, in which the intervention factor (psychedelic drug, placebo) and instructional set provided to each participant (receiving psychedelic drug, receiving placebo) are systematically crossed with each other. This design offers a potentially rigorous experimental means for separating pharmacological effects of the drug from participant expectations but is most suitable for mechanistic studies of acute drug effects, rather than clinical trials examining treatment efficacy. To date, there are no published reports of this design being used in psychedelic drug research, possibly because of its high costs (Schenberg 2021). Although researchers have begun to address the methodological challenges associated with masking, treatment expectations, and their combined impact that can bias study results, there is a need to advance the rigor of future research. We build upon this work in the next section by elaborating on recommendations for improving psychedelic clinical trials.

**Novel recommendations to improve future research** Experimental confounds related to expectancies and placebo effects in psychedelic studies largely stem from inadequate masking. Therefore, our recommendations are primarily focused on how to improve masking in psychedelic trials through a combination of procedures intended to decrease participants’



**Fig. 3** Recommendations for improving methodology in psychedelic trials. Overview of our recommendations for improving experimental methodology in future clinical trials with psychedelics

confidence in their assigned treatment arm (Fig. 3). As our review of others' pioneering work makes clear, adequate masking involves critical decision points at every step in the lifecycle of a clinical study. Our suggestions follow suit, noting elements for consideration in study development and design, participant recruitment and selection, outcomes and endpoints, study procedures, and analysis plans. It should be noted that masking is not an all-or-nothing phenomenon; incorporating a portion of these suggestions can incrementally reduce participants' confidence in their treatment arm assignment and thereby attenuate the influence of treatment-nonspecific factors in interpretations of clinical trials.

### Study development and design

The choice of a control condition, the number of study arms, and overall design should be determined by the specific purpose of the study (Freedland, 2020; Gold et al., 2017). For example, although an open-label study design does not mask participants or control for treatment-nonspecific factors, it may be appropriate when the purpose of the study is to examine safety, feasibility, or proof-of-concept. If the purpose is to examine treatment efficacy, inactive control conditions (e.g., treatment-as-usual, waitlist controls) should be included at the minimum to control for some treatment-nonspecific factors, such as natural history or regression to the mean. A stronger study design to test for efficacy would include an active control condition, such as an active placebo that mimics some of the acute effects of a psychedelic. Including both an active and inactive control condition (i.e.,

3-arm design) is a promising way to disentangle placebo effects (Fillingim and Price 2005; Smith et al. 2020; Vase and WartolowVaseska 2019), because 3-arm trial designs allow for comparisons between both the treatment and the active placebo conditions with the inactive control condition to delineate treatment-specific effects from placebo effects (see Fig. 1a). There are also alternative study designs that may be especially useful because of psychedelic trials' vulnerability to large placebo effects. Sequential parallel designs with a placebo run-in period can reduce the size of placebo effects by excluding "placebo responders" from the subsequent treatment phase (Campbell et al. 2019; Dworkin et al., 2010; Ivanova et al. 2016; Tamura and Huang 2007). This alternative design can be implemented in psychedelic trials by giving all participants an active placebo in the first phase and then randomly assigning only the participants who did not respond to the initial treatment (i.e., placebo nonresponders) to the psychedelic or placebo in the second phase. This placebo run-in period creates a subgroup for analysis that increases the sensitivity to detect a treatment-specific effect (Dworkin et al., 2010; Ivanova et al., 2016); however, a recent systematic review challenges the notion that this design actually reduces the measured placebo response (Scott et al. 2021).

We also recommend designing studies with a single psychedelic administration when possible, given our current understanding regarding the efficacy of psychedelic therapy. There are compelling reasons to believe that multiple psychedelic dosing sessions may have therapeutic advantages (Bousso et al. 2013; Leger and Unterwald 2021;

Mithoefer et al. 2019), and this treatment model is very likely to be adopted in clinical practice if these therapies become FDA-approved. On the other hand, the current controversies surrounding psychedelic therapy are focused on whether there is any drug-specific benefit of the complex therapeutic intervention. The answer to this basic question is very likely to inform regulatory decisions, cost-effectiveness models, and coverage by insurers, and is dependent on adequately masked trials. To that end, studies with only a single dosing session are likely to be superior in supporting adequate masking compared to studies with multiple dosing sessions. That is, once participants have experienced the subjective effects of a substance, they are more likely to identify that substance if it is readministered or recognize that a different substance has been given, compromising the conclusions that can be drawn from the trial (Wilsey et al. 2016). Therefore, we recommend between-subjects designs with a single dosing session when evaluating treatment efficacy.

Several trials have included an open-label crossover component, wherein patients assigned to the inactive control arm are offered the opportunity to receive open-label psychedelic therapy after completing the final post-treatment assessment (Wolfson et al. 2020). Some have argued that this design feature is ethically mandatory in order to provide the patient with the best possible chance of therapeutic response. We disagree with the idea that the standard of care, or optimal care, involves offering unregulated and unapproved psychedelic therapy, particularly when the goal of these trials is to establish the efficacy of these same interventions. We recommend incorporating well-established strategies to minimize harm to participants that may arise if an experimental therapy is either harmful, or conversely highly effective, rendering placebo treatment unethical. “Stopping rules” are predefined time points where an interim analysis for efficacy can be performed to identify these situations and minimize harm. Alternatively, adaptive randomization based on outcome (see below) can achieve a similar goal while maintaining statistical power (Dragalin 2011)<sup>3</sup>. We also emphasize the importance of including robust psychotherapeutic support in any treatment arm when dealing with high-risk populations selected for treatment resistance, both to maximize patient safety and monitoring and to better assess drug-specific enhancement of psychotherapy as discussed previously.

<sup>3</sup> These strategies are complementary to existing mechanisms for patients to try unapproved therapies, instituted as the Right to Try Act in the USA, as well as expanded access clinical programs (Holbein et al. 2015)

## Participant recruitment and selection

We recommend recruiting psychedelic-naive participants when possible for clinical trials. Masking an individual’s treatment condition is much more feasible if they have no prior experience with that substance and are less certain about what effects to expect (i.e., process expectations; Tambling 2012; Wilsey et al. 2016). On that basis, participants should be naive to the active placebo as well. Ostensibly, psychedelic-naive individuals would have less confidence as to whether they received the treatment or active placebo, particularly if the active placebo had hallucinogenic effects. Carbonaro et al. (2018) demonstrated that experienced hallucinogen users are highly accurate at differentiating between whether they received psilocybin or DXM, but those without prior hallucinogen use may be easier to convince, especially if this strategy is combined with other recommendations given here (e.g., incomplete disclosure of study design, between-subjects designs with a single drug administration). It should be noted, however, that a challenge with this design is that several psychoactive substances (e.g., cannabis, opioids) are known to elicit different subjective and behavioral responses in drug-naive individuals compared to those with past experience (Solowij et al. 2019). This appears to be the case with psychedelics too, as demonstrated by a negative relationship between number of previous psychedelic uses and the intensity of acute effects (Aday et al. 2021). Thus, the phenomenological experience and intensity of drug effects may differ in first-time users, which could limit generalizability. If recruiting only psychedelic-naive participants is not feasible given the increasing number of recreational users (Yockey et al. 2020), then imposing clear exclusion criteria, such as restrictions on number of lifetime uses or use within the past 12 months, should be incorporated.

## Outcomes, assessments, and endpoints

The choice of outcomes, assessments, and endpoints can have a large impact on the evaluation of treatment benefit and overall methodological rigor of psychedelic clinical trials. The primary endpoint for a trial should be well-defined, reliable, and represent a clinically meaningful outcome of how a patient feels, functions, or survives (e.g., Fleming and Powers 2012; US FDA 2009). Outcome measures should be consistent with expert recommendations or consensus statements for a given disease or condition under study when available (e.g., Deyo et al. 2014), and the minimal clinically important difference in the primary outcome measure that represents a treatment benefit should be set a priori (e.g., Dworkin et al. 2008, 2009). There are unresolved questions regarding the long-term efficacy of psychedelic therapy. Lasting, clinically significant improvements following

psychedelic therapy, regardless of any placebo group difference, are likely more important to patients, providers, and stakeholders than an acute improvement that is not maintained. However, given the current level of evidence and controversy regarding the drug-specific efficacy of the treatment, we emphasize the primary importance of rigorous, well-controlled trials is to define clear evidence of benefit that outlasts the acute drug effect. The specific timing of outcomes will depend heavily on the indication under consideration. Although long-term follow-ups provide a more complete understanding of treatment effects, especially in trials on chronic conditions, they are still susceptible to placebo effects and selection bias affecting trials from the outset. For example, a well-designed, masked RCT showed that arthroscopic knee surgery was never better than placebo surgery across 2 years of assessments (Moseley et al. 2002).

We recommend using multiple methods of measurement to comprehensively examine the effects of psychedelic therapy in clinical trials. Patient-reported outcomes (PROs) assess the status of a patient's health condition (e.g., disease symptoms, functioning) directly from the patient and are commonly used as endpoints in clinical trials (Mercieca-Bebber et al. 2018; US FDA 2009). Including valid, reliable, and clinically informative PRO measures is valuable because they capture patient-centered perceptions of meaningful change and have downstream influence on clinical decision-making, drug labeling claims, and health policy (Calvert et al. 2018; Doward et al. 2010). Clinician-administered assessments or observer reports can also be useful in psychedelic trials as they avoid potential self-report biases of PROs; however, these types of assessments are also vulnerable to methodological issues, such as low interrater reliability and rater bias (Kobak et al. 2007). Therefore, when feasible, trials should also include objective and reliable measures, such as biomarkers and/or behavioral tasks that reflect component processes related to the index pathology. Two categories of biomarkers recognized by the FDA (Smith et al. 2017; US FDA 2020) that may be particularly relevant for psychedelic clinical trials are predictive biomarkers and surrogate endpoints. Predictive biomarkers indicate whether certain participants respond differentially to the treatment or placebo and can be used to stratify randomization on variables of interest that may maximize the efficiency of a trial and minimize the risk of exposing additional patients to an unproven treatment (Strimbu and Tavel 2010). Surrogate endpoint biomarkers include accurate and well-validated lab measures or physical signs that reliably predict or stand in for a clinically meaningful endpoint (e.g., biomarkers of abstinence; Johnson et al. 2014; Fleming and Powers 2012). Not all diseases or health conditions have biomarkers that predict treatment benefit or represent clinical endpoints, but when available, inclusion of these types of biomarkers may lead to more efficient trials with less bias (Fleming and

Powers 2012). Because psychedelic clinical trials are particularly expensive, one must weigh the tradeoffs between trial costs and participant burden with the addition of biomarkers, long-term follow-ups, and lengthy assessments.

### **Study procedures: managing and measuring treatment expectations**

Several pragmatic steps can be taken at the beginning stages of a study to manage participants' expectation bias. We do not currently have sufficient data to claim that psychedelic therapy is an effective treatment; therefore, investigators should emphasize the uncertainty regarding the treatment efficacy, rather than insinuating that the treatment will improve participants' symptoms (Erpelding et al. 2020; Evans et al. 2021; Gewandter et al. 2020; Smith et al. 2020). This communication about the uncertainty of treatment efficacy should be consistent across recruitment materials, initial contact with potential participants, consent forms, and any interactions with participants. Moreover, in trials comparing psychedelic therapy to placebo, drug effects should be explained neutrally (Smart et al. 1966). For example, participants can truthfully be informed about possible drug effects while also noting that there is significant variability between people—some people have strong reactions to a psychedelic while others have very mild reactions (Griffiths et al. 2016). Similarly, in studies in which both treatment arms receive psychotherapy, the investigator can honestly describe psychotherapy as an effective treatment whether or not it is paired with a psychedelic. To ensure this clinical equipoise and manage participants' expectations, all study staff should be masked to treatment arm assignment and trained to present the study and arms of the trial neutrally.

In addition to managing expectations, it is important to measure participants' treatment expectations. We and others (e.g., Muthukumaraswamy et al. 2021) recommend the use of established measures of expectancy, such as the Stanford Expectations of Treatment Scale (Younger et al. 2012), which is a valid and reliable measure of participants' positive and negative treatment expectancies. The scale includes six items that can easily be adapted across research contexts to identify differences in expectancies between treatment groups as well as relationships between treatment expectancies and outcomes. The Credibility and Expectancy Questionnaire (Deville and Borkovec 2000) can also be used to measure the degree to which a participant thinks and feels the treatment will improve their symptoms or functioning. Furthermore, several face-valid questions, such as “how helpful do you believe the treatment will be for improving your [primary symptom]?”, have been used successfully to measure treatment expectations in previous research (e.g., Sherman et al. 2010). Another option is to conduct semi-structured interviews, possibly during

participant preparation and integration sessions, and use qualitative analyses to assess participants' positive and negative treatment expectations (e.g., Eaves et al. 2015). Because of the aforementioned issues with unmasking following a psychedelic session, and the interaction between masking and expectations, it may be useful to measure treatment expectations after the drug dosing session in addition to those at baseline. Arguably, expectations at baseline may be predictive of subjective effects during the psychedelic session, and expectations at post-session may be predictive of changes in clinical outcomes. This speculation remains to be tested, but it is worthwhile to systematically evaluate the natural dynamics of expectations during psychedelic trials and examine whether expectations change after the dosing session.

### Study procedures: incomplete disclosure

We have reviewed studies where incomplete disclosure has been used to reduce participants' certainty regarding their treatment assignment (Bershad et al. 2019; Carbonaro et al. 2018; Griffiths et al. 2006; Reissig et al. 2012). In designing a trial, it is critically important to distinguish "incomplete disclosure" from "deception." Most institutional review boards have internally defined these respective procedures; however, "deception" is generally agreed to mean that the investigators provide false information to a participant whereas "incomplete disclosure" indicates that the subject is not fully informed about the purpose or design of the study. These strategies are controversial—the ethics of omitting important information about a study and misleading participants is an area of ongoing debate (Miller et al. 2005; Roulet et al. 2017). Implementing any deceptive practice requires thorough scientific justification and authorization by institutional review boards. Empirical evidence in healthy adults suggests that research participants may not be adversely affected by deception (Mundt et al. 2017); however, in the context of clinical trials in which therapeutic alliance is critical for patient safety and treatment efficacy, deception may be particularly ill-advised. If it is considered ethically appropriate, though, withholding information from participants as well as study staff about the number of study arms and the exact doses administered may be particularly effective for enhancing masking success. Providing a vague, incomplete description of the study structure and a range of possible dosages may be best suited for standard, two-armed RCT designs (and avoids the need to use an alternative study design that requires a significantly larger sample size for adequate statistical power). Without the cues of knowing that it is only possible to receive the experimental treatment or placebo (e.g., a high dose or an ultra-low dose of a psychedelic), it may be difficult for both the participant and staff to develop a firm belief about the participant's

treatment condition. Similarly, listing the side effects of all of the potential study drugs together—instead of listing effects specific to each substance—may be an ancillary strategy to reduce participants' confidence in their treatment arm assignment while still fully informing them of all the drug effects they may be exposed to (Boutron et al. 2006). In a recent study with 5-MeO-DMT, researchers withheld the identity of the study drug but informed participants that they would be receiving a tryptamine psychedelic (Reckweg et al. 2021); this may be a useful method for managing expectations in cases where participants could have distinct expectations regarding specific psychedelic substances. A related recommendation to improve methodological rigor in the field is for researchers to report what drug effects participants were informed about prior to the study.

Incomplete disclosure to participants and study personnel regarding key elements of a study's design may help to meet a central objective of masking: establishing "a state of ambivalence" about treatment allocation to minimize the impact of beliefs on study outcomes (Mathieu et al. 2014). Ensuring that study staff receive the same information as participants and remain unaware of the true design throughout the study is critical, as feedback from observers is known to influence participants' clinical outcomes (Colagiuri and Boakes 2010; Hróbjartsson et al. 2012). It is important to acknowledge that undertaking this effort—concealing fundamentals of study design from staff as well as participants—is challenging from a practical standpoint, requiring careful management of access to information about the study (e.g., a "cone of silence"). Using incomplete disclosure or deception also necessitates appropriate debriefing protocols, as well as development of masking assessments that avoid revealing the true study design. Most assessment tools in the clinical trial literature measure perceived treatment assignment as nominal data and implicitly indicate study design (i.e., "Do you think you received the active treatment or placebo?"). Probing participants' and staff members' beliefs using ordinal/parametric scales may not only allow investigators to maintain uncertainty about the design, but also has the advantage of increasing statistical power (Laferton et al. 2017).

### Study procedures: active placebo

Use of an active placebo has a clear rationale for psychopharmacology studies. However, as reviewed above, efforts to mask the unique subjective effects of psychedelics have had limited success. Our choices are largely constrained by a limited understanding of how psychedelics produce therapeutic benefits. For example, a drug that mimics psychedelic effects but provides no therapeutic benefit could potentially be an excellent active placebo. However, the internal contradiction in this strategy becomes apparent if, as several

researchers argue (Yaden and Griffiths 2020), the subjective effects produced by psychedelics (particularly mystical states) themselves drive therapeutic benefit. Although intuitive, this hypothesis is nonetheless unproven and a thorough evaluation is beyond the scope of this review; we instead refer the reader to an excellent summary of arguments for and against this idea (Olson 2020; Yaden and Griffiths 2020). We anticipate that future research will clarify whether mystical states induced by means other than psychedelics such as hypnosis (Lynn and Evans 2017), holotropic breathwork (Puente 2014), meditation (Russ and Elliott 2017), virtual reality (Glowacki et al. 2020), or non-psychedelic psychoactive drugs (Earleywine et al. 2021) are sufficient for therapeutic effects observed in psychedelic therapy trials, such as smoking cessation and symptomatic relief from depression in appropriate target populations.

A deeper understanding of the neural systems and neurochemistry required for psychedelics' therapeutic effects may lead to highly effective comparators for use in clinical trials. A recent clinical study investigating the antidepressant mechanism of ketamine illustrates that the acute subjective effects of a psychedelic-class drug may be separable from its therapeutic effects. Williams et al. (2018, 2019) found that a high dose of an opioid antagonist, naltrexone, effectively blocked ketamine's antidepressant and anti-suicidal effects but had a minimal impact on ratings of ketamine-induced dissociation. This small study was met with some controversy (Heifets et al. 2019; Marton et al. 2019; Yoon et al. 2019) and requires replication in a larger independent sample. Also, notably, the authors did not formally assess masking efficacy in the respective treatment conditions. Nonetheless, these findings suggest a powerful active placebo comparator for future studies of ketamine, and potentially other psychedelics. Similarly, for classical psychedelics like psilocybin, pharmacological agents may be discovered that interrupt neuroplastic processes triggered by psilocybin, but do not interfere with its acute psychedelic effects. Another highly innovative approach in development (NCT04842045) pairs psilocybin with an amnestic drug (midazolam, a benzodiazepine). This study is focused on safety. The broader hypothesis, yet to be tested, is that psychedelic and mystical states evoked in participants who do not form memories of the experience are not therapeutic, likely because participants' amnesia prevents subsequent therapeutic integration of the psychedelic experience. An alternate outcome may be that participants do experience therapeutic benefit, but are effectively masked to their assigned treatment condition by virtue of midazolam-induced amnesia. In this case, a near-perfectly controlled, masked study design is achieved, with an easily interpretable finding for psilocybin's efficacy, uncomplicated by differential placebo or nocebo effects in patients receiving midazolam alone versus midazolam plus psilocybin. We eagerly anticipate results from this

pioneering line of inquiry and note several challenges. In addition to the ethical considerations of using amnestic agents in psychiatric populations, there are technical considerations that may confound this approach, including uncertainty as to whether midazolam retains its amnestic property when paired with a psychedelic, whether amnestic doses of midazolam produce a degree of sedation that precludes entry into a mystical state, or whether midazolam directly blocks therapeutic psychological or neural mechanisms induced by psychedelic medications.

Psychedelic therapy may be an uninterrupted whole, requiring the drug, psychedelic experience, and associated psychotherapy to achieve any therapeutic benefits (Sessa 2014). In this case, which should be assumed true until proven otherwise, there is still a pragmatic need to identify pharmacological and somatic placebo treatments that adequately mask psychedelic effects. Although we have no evidentiary basis to recommend specific active placebos beyond those that have been attempted, substances with hallucinatory effects (e.g., ketamine, DXM, and high doses of tetrahydrocannabinol) may be compelling options, especially when combined with drug-naïve participants. We strongly support studies specifically devoted to developing and testing active placebos for use in therapeutic clinical trials. The need to develop active placebos for participants with past psychedelic use is particularly important given the likely decrease in psychedelic-naïve participants that can be recruited for clinical therapeutic studies in the coming years.

Design of an active placebo ought to be considered in concert with other study design elements described above, with the overarching goal of reducing a prospective study participant's certainty of their treatment condition. For example, if testing psilocybin's efficacy for major depressive disorder, investigators may combine active placebo and incomplete disclosure to balance expectancy effects across treatment arms. For simplicity, the study could be designed as a two-arm comparison of high-dose psilocybin versus ultra-low-dose (ineffective) psilocybin plus an active placebo. During the informed consent process, participants would truthfully be informed that they will receive a range of psilocybin doses and may also receive an active placebo, with full disclosure that the purpose of the active placebo is to reduce their certainty of treatment assignment. The number of study arms (two, in fact) and the likelihood that their assigned psilocybin dose would be effectively non-therapeutic would not be disclosed. Furthermore, informed consent could include information that subthreshold (but not ultra-low) psilocybin may have therapeutic value, although, again, it would not be disclosed that no participants would be assigned to a subthreshold dose group. In this case, the specific goal of an active placebo might be to mimic aspects of a high-dose psilocybin dose, which could be achieved with DXM or perhaps a combination of a benzodiazepine

and a mild stimulant. Taken together, participants would be informed of all the possible treatment conditions and may be reasonably uncertain as to whether they received a high therapeutic dose of psilocybin versus an ultra-low dose plus active placebo.

### **Analysis: assessing and reporting outcomes related to trial design**

The set of treatment-nonspecific effects, collectively termed “the placebo effect,” and effective masking are key considerations for designing an interpretable study involving psychoactive drugs. Anticipating the placebo effect, measuring the contribution of expectancies, assessing the effectiveness of masking, and systematically reporting these data will set standards and lead to iterative improvements in trial design. These factors ought to be considered at every step in the lifecycle of a clinical study. We specifically recommend calculating statistical power based on known placebo effect sizes, obtaining repeat baseline measures of the primary outcome(s), measuring expectancies and masking success, and analyzing primary outcomes using expectancy and perceived (rather than actual) treatment arm as covariates.

Estimating the size of the placebo effect informs statistical power calculations, which, if resources are limited, may impact the feasible number of treatment arms. A common method of estimating the size of the placebo effect in a trial is to compare outcomes in the placebo arm to a “no treatment” arm (Hróbjartsson and Gøtzsche 2010; Wampold et al. 2016). However, given the previously discussed “hype” around psychedelics, participants randomly assigned to the “no treatment” arm would likely experience disappointment and nocebo effects from their knowledge of not being in the active treatment. An alternative method of partitioning the placebo effect from the treatment effect may be to compare against a “placebo benchmark” (Jones et al. 2021). Jones and colleagues found that the effect size of the placebo effect was uniform across different treatment approaches for depression (pooled Hedge’s  $g = 1.05$ ). In areas where the size of the placebo effect has been well-established, researchers may be able to compare their anticipated effect size against a criterion. Investigators can also take simple steps to minimize some components of the placebo effect, such as regression to the mean. We recommend that investigators perform repeat baseline assessment of their outcome of interest and only enroll participants with stable response characteristics. This procedure may be more cost-effective than including an untreated control condition to estimate regression to the mean.

We strongly recommend measuring the factors that make up the placebo effect. Prior to conducting any study procedures (e.g., preparation sessions), participants’ treatment expectations should be measured as described above.

Measuring masking efficacy is similarly important and should be appropriately timed. In many cases, the clinical benefits of psychedelics may be rapid (Majić et al. 2015; Murphy-Beiner and Soar 2020). We recommend measuring participant- and therapist-perceived treatment allocation, certainty of treatment allocation, and the reason for their guess both immediately after the psychedelic dosing session(s) and at the end of the study.

Including two measurement occasions may help determine whether participants and therapists guessed the treatment allocation based on the subjective effects during the treatment session or from changes in clinical symptoms over time (Katz, 2021; Kolahi et al. 2009). We agree with Katz (2021) that accurate guesses of treatment allocation due to treatment efficacy should not be considered unmasking. To further redress the influence of masking, we suggest using clinical assessors who are unaware of the study design and participant treatment allocation to collect all relevant measures. Clinical assessors should also be asked about perceived participant treatment allocation at the end of the study (Katz 2021). We again emphasize that investigators should create protocols and adherence plans for all relevant study staff to maximize the chances that masking is maintained throughout the study.

Participant expectations and functional unmasking may be unavoidable sources of bias that impact internal validity and the inferences that can be drawn from study results (Higgins et al., 2011; Kolahi et al., 2009). However, modern adaptive trial designs can help investigators at least achieve an even distribution of these biases across conditions. A thorough discussion of adaptive designs is beyond the scope of this review, and we refer the reader to two useful summaries, including draft guidance from the FDA on adaptive trial design for industry (FDA 2019; Pallmann et al. 2018). In short, investigators may consider using expectancy and participant-assessed treatment conditions to create balanced randomization blocks (i.e., covariate-adaptive treatment assignment) just as other clinical trials stratify recruitment on the prevalence of comorbidities, sex, and other factors that may differentially impact treatment outcomes. For small exploratory trials, it may not be possible to balance on multiple pre-treatment variables; therefore, the decision to balance recruitment on treatment outcome expectations must be weighed against other recruitment priorities.

A major benefit of measuring expectancies and masking efficacy is that these factors can be used as covariates in the analysis of primary study outcomes, and the specific effects of expectancy and treatment arm guess on outcome can be evaluated. In the previously discussed microdosing study by van Elk et al. (2021), researchers initially found that microdoses of psilocybin led to greater ratings of awe than placebo; however, after adding baseline expectations as a covariate to the analyses, the difference between conditions

was non-significant. In a study that employs an effective active placebo, outcomes can be analyzed according to the drug that participants think they received compared to the drug they actually received. In a study measuring pleasantness of affective touch, Bershad et al. (2019) found a significant effect of MDMA compared to an active placebo, methamphetamine. A substantial number of participants who received methamphetamine believed they had received MDMA (38.9%). Analyzing outcomes using a participant's guess as a covariate showed no effect in this latter group. This comparison strongly reinforced the authors' conclusion that the effect of MDMA on affective touch was drug-specific and not a product of participants' expectations.

## Beyond the scope of RCTs

One notion to consider is embracing expectancy and placebo effects. The important role of expectancies in psychedelic therapy blurs the line between treatment-specific and treatment-nonspecific effects and raises the broader question: rather than eliminating treatment-nonspecific effects, should trialists be looking for ways to optimize and synergize them with treatment interventions to enhance clinical outcomes (Colloca and Barsky 2020; Enck et al. 2013)? Although no formalized manual exists on how to boost expectancy in psychotherapy, inducing positive expectations has been shown to enhance the effectiveness of a variety of health interventions (Bingel et al. 2011; Flowers et al. 2018; Kaptchuk et al. 2020), a strategy which could seemingly be tailored to—and be particularly synergistic with—psychedelic treatments as well. As discussed previously, placebo and drug-specific effects are likely to be interactive rather than additive (Kube and Rief 2017). Thus, it may be the case that the “therapeutic window” opened by psychedelics is an emergent property of a complex system comprising expectations, drug effects, setting, and therapeutic alliance. It may be impossible to isolate an individual component of this complex package in an RCT. Critically, this does not condemn psychedelic therapy as being no more effective than placebo, but means that the current gold standard clinical trial design may not be sensitive to detecting the therapeutic effect of an individual treatment element.

A potential solution to this dilemma may be to shift focus from efficacy trials and the use of explanatory or confirmatory RCT designs towards pragmatic clinical trial designs (PCTs) that have an alternative goal of assessing treatment effectiveness. Whereas internal validity (i.e., objective comparison of drug vs placebo in tightly controlled settings with homogenous groups) is the major objective of an explanatory or confirmatory trial, external validity and the generalizability of treatment effectiveness are the primary focus of a well-designed PCT. Consequently, PCTs offer potential

“real-world” tests of clinical effectiveness and the generalizability of outcome data, rather than isolation of the active ingredient for change. To achieve these goals, PCTs typically include one or more alternative therapies to the treatment under study, rather than active or inactive placebos, and participants are normally recruited from a broad “real-life” clinical population, with few exclusions or restrictions on participation. Although pragmatic trials are normally conducted in the fourth, post-marketing phase of drug development, Carhart-Harris et al. (2021) have argued cogently for the potential benefits of pragmatic designs being used earlier to broadly assess the clinical effectiveness of current psychedelic treatments, either as an alternative or complement to the much narrower focus of current RCTs.

Lastly, a closely related approach to consider when testing the effectiveness of psychedelic therapy is to evaluate large-scale population data using so-called “natural experiments.” Natural experiments provide an alternative to RCTs by taking advantage of circumstances whereby naturally occurring events can be linked to variables of interest (Thapar and Rutter 2019). This type of design is necessary when randomly assigning individuals to masked conditions is not possible because of ethical or logistical constraints, such as when studying maltreatment or child neglect (Rutter 2007). If the challenges related to expectations and masking with psychedelics preclude rigorous RCTs, natural experiments may be another method of evaluating the treatment's effects. With the recent legalization of psilocybin therapy in Oregon as well as successful decriminalization movements across the USA (Aday et al. 2020a; Marks and Cohen 2021), it is possible that objective indices related to mental health (e.g., suicide rates, emergency room visits for psychiatric issues) could precipitously decrease at the population level if psychedelics are indeed an effective treatment for a variety of psychiatric conditions. Although it is unclear what the initial accessibility of these treatments will be to individuals in states such as Oregon (Williams and Labate 2020), if positive trends in mental health are observed at the population level after the introduction of legal psychedelic therapy, the role of expectations may be considered immaterial to the broader benefits to society.

## Conclusion

Accurate detection of treatment-specific effects in clinical trials is an intrinsically complex task across areas of research as study personnel and participant expectations interact dynamically with masking and therapeutic outcomes. Psychedelic studies are particularly challenging as they must address additional confounds related to “hype” and salient psychoactive effects that hinder treatment arm masking to an extensive degree. On one hand, to characterize clinical



efficacy and safety, it is an essential challenge for the field to separate pharmacological effects from multiple, interactive socio-psychological influences in psychedelic medicine. Innovative, disruptive experimental designs may be needed to this end. On the other hand, at a practical level, it is important from a public health standpoint to identify methods of optimizing psychedelic treatment outcomes, perhaps by utilizing expectancies. These results could potentially guide clinical decision-making.

Traditional placebo masking with inert comparators is insufficient for high-dose psychedelic studies, and this review highlights that this issue often extends to psychotherapy and pharmacology research more broadly. Here, recommendations are presented for improving the methodological rigor of future psychedelic studies that addresses issues related to expectations and participant masking. Specifically, we provide guidelines on study design (e.g., incomplete disclosure of treatment arms, neutral explanation of drug effects), participant recruitment and selection (e.g., include psychedelic- and active placebo-naïve participants), outcomes and endpoints (e.g., include biomarkers and behavioral measures), control conditions (e.g., use active comparators), and analyses (e.g., test masking efficacy, control for pre-treatment expectations, compare against placebo benchmark). Although these recommendations are tailored to psychedelic studies, they can be incorporated into psychotherapy and pharmacology research more broadly to increase precision in identifying treatment-specific effects. Doing so may improve methodological rigor and identification of effective interventions across areas of medicine.

**Acknowledgements** We would like to thank members of the Translational Psychedelic Research (TrPR) Program at the University of California, San Francisco (UCSF) for providing comments on this manuscript.

## Declarations

**Conflict of interest** JDW was a paid consultant for Silo Pharma and Filament Health last in June 2021. BDH is a paid consultant for Clairvoyant Therapeutics. None of the other co-authors have any conflicts of interest to declare.

## References

- Aday JS (2021) Delineating the association between psychedelic drug use and facets of gratitude, aesthetic experience, and relationships with nature. Dissertation, Central Michigan University
- Aday JS, Davis AK, Mitzkovitz CM, Bloesch EK, Davoli CC (2021) Predicting reactions to psychedelic drugs: a systematic review of states and traits related to acute drug effects. *ACS Pharm Trans Sci* 4:424–435. <https://doi.org/10.1021/acspsci.1c00014>
- Aday JS, Bloesch EK, Davoli CC (2020a) 2019: A year of expansion in psychedelic research, industry, and deregulation. *Drug Sci Pol Law* 6. <https://doi.org/10.1177/2050324520974484>
- Aday JS, Mitzkovitz CM, Bloesch EK, Davoli CC, Davis AK (2020b) Long-term effects of psychedelic drugs: a systematic review. *Neurosci Biobehav Rev* 113:179–189. <https://doi.org/10.1016/j.neubiorev.2020.03.017>
- Atlas LY (2021) A social affective neuroscience lens on placebo analgesia. *Trends Cogn Sci*. <https://doi.org/10.1016/j.tics.2021.07.016>
- Barnett AG, Van Der Pols JC, Dobson AJ (2005) Regression to the mean: what it is and how to deal with it. *Int J Epidemiol* 34:215–220. <https://doi.org/10.1093/ije/dyh299>
- Barsky AJ, Klerman GL (1983) Overview: hypochondriasis, bodily complaints, and somatic styles. *Am J Psychiatry* 140:273–283. <https://doi.org/10.1176/ajp.140.3.273>
- Basoglu M, Marks I, Livanou M, Swinson R (1997) Double-blindness procedures, rater blindness, and ratings of outcome: observations from a controlled trial. *Arch Gen Psychiatry* 54:744–748. <https://doi.org/10.1001/archpsyc.1997.01830200078011>
- Bedi G, Hyman D, de Wit H (2010) Is ecstasy an “empathogen”? Effects of  $\pm$ 3, 4-methylenedioxymethamphetamine on prosocial feelings and identification of emotional states in others. *Biol Psychiatry* 68:1134–1140. <https://doi.org/10.1016/j.biopsych.2010.08.003>
- Bershad AK, Mayo LM, Van Hedger K, McGlone F, Walker SC, de Wit H (2019) Effects of MDMA on attention to positive social cues and pleasantness of affective touch. *Neuropsychopharm* 44:1698–1705. <https://doi.org/10.1038/s41386-019-0402-z>
- Bingel U, Wanigasekera V, Wiech K, Mhuirheartaigh RN, Lee MC, Ploner M, Tracey I (2011) The effect of treatment expectation on drug efficacy: imaging the analgesic benefit of the opioid remifentanyl. *Sci Transl Med* 3:70ra14–70ra14. <https://doi.org/10.1126/scitranslmed.3001244>
- Bishop MD, Mintken P, Bialosky JE, Cleland JA (2019) Factors shaping expectations for complete relief from symptoms during rehabilitation for patients with spine pain. *Physiother Theory Prac* 35:70–79. <https://doi.org/10.1080/09593985.2018.1440676>
- Bland JM, Altman DG (1994) Statistics notes: some examples of regression towards the mean. *BMJ* 309:780. <https://doi.org/10.1136/bmj.309.6957.780>
- Bouso JC, Fábregas JM, Antonijuan RM, Rodríguez-Fornells A, Riba J (2013) Acute effects of ayahuasca on neuropsychological performance: differences in executive function between experienced and occasional users. *Psychopharm* 230:415–424. <https://doi.org/10.1007/s00213-013-3167-9>
- Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hróbjartsson A, Ravaut P (2006) Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med* 3:e425. <https://doi.org/10.1371/journal.pmed.0030425>
- Burke MJ, Blumberger DM (2021) Caution at psychiatry’s psychedelic frontier. *Nat Med* 27(10):1687–1688
- Burns JW, Jensen MP, Thorn BE, Lillis TA, Carmody J, Newman AK, Keefe F (2021) Cognitive therapy, mindfulness-based stress reduction, and behavior therapy for the treatment of chronic pain: a single-blind randomized controlled trial. *Pain*. <https://doi.org/10.1097/j.pain.0000000000002357>
- Calvert M, Kyte D, Mercieca-Bebber R et al (2018) Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *JAMA* 319:483–494. <https://doi.org/10.1001/jama.2017.21903>
- Campbell CM, Gilron I, Doshi T, Raja S (2019) Designing and conducting proof-of-concept chronic pain analgesic clinical trials. *Pain Rep* 4. <https://doi.org/10.1097/PR9.0000000000000697>
- Carbonaro TM, Johnson MW, Hurwitz E, Griffiths RR (2018) Double-blind comparison of the two hallucinogens psilocybin and dextromethorphan: similarities and differences in subjective

- experiences. *Psychopharm* 235:521–534. <https://doi.org/10.1007/s00213-017-4769-4>
- Carhart-Harris RL, Kaelin M, Whalley MG, Bolstridge M, Feilding A, Nutt DJ (2015) LSD enhances suggestibility in healthy volunteers. *Psychopharm* 232:785–794. <https://doi.org/10.1007/s00213-014-3714-z>
- Carhart-Harris RL, Wagner AC, Agrawal M et al (2021) Can pragmatic research, real-world data and digital technologies aid the development of psychedelic medicine? *J Psychopharmacol*. <https://doi.org/10.1177/02698811211008567>
- Carpenter DE (2020) 5-MeO-DMT: the 20-minute psychoactive toad experience that's transforming lives. *Forbes*. <https://www.forbes.com/sites/davidcarpenter/2020/02/02/5-meo-dmt-the-20-minute-psychoactive-toad-experience-thats-transforming-lives/?sh=293db2d38a1>
- Colagiuri B, Boakes RA (2010) Perceived treatment, feedback, and placebo effects in double-blind RCTs: an experimental analysis. *Psychopharm* 208:433–441. <https://doi.org/10.1007/s00213-009-1743-9>
- Colloca L, Barsky AJ (2020) Placebo and nocebo effects. *New Eng J Med* 382:554–561. <https://doi.org/10.1056/NEJMr1907805>
- Constantino MJ, Arnkoff DB, Glass CR, Ametrano RM, Smith JZ (2011) Expectations *J Clin Psych*, 67: 184–192. <https://doi.org/10.1002/jclp.20754>
- Constantino MJ, Vısla A, Coyne AE, Boswell JF (2018) A meta-analysis of the association between patients' early treatment outcome expectation and their posttreatment outcomes. *Psychother* 55:473–485. <https://doi.org/10.1037/pst0000169>
- Cunningham JA, Kyprı K, McCambridge J (2013) Exploratory randomized controlled trial evaluating the impact of a waiting list control design. *BMC Med Res Methodol* 13:1–7. <https://doi.org/10.1186/1471-2288-13-150>
- Cusack CM (2020) Anne Hamilton-Byrne and the family: charisma, criminality, and media in the construction of an Australian “cult” leader. *J Alt Emerg Rel* 24:31–54. <https://doi.org/10.1525/nr.2020.24.1.31>
- D'Onofrio G, Fiellin DA, Pantalon MV et al (2012) A brief intervention reduces hazardous and harmful drinking in emergency department patients. *Ann Emerg Med* 60:181–192
- Davis AK, Barrett FS, May DG et al (2021) Effects of psilocybin-assisted therapy on major depressive disorder: a randomized clinical trial. *JAMA Psych* 78:481–489. <https://doi.org/10.1001/jamapsychiatry.2020.3285>
- de Wit H, Bershad AK, Hutchison W et al (2021) Can MDMA change sociopolitical values? Insights from a research participant. *Biol Psychiatry* 89:e61–e62. <https://doi.org/10.1016/j.biopsych.2021.01.016>
- Devilley GJ, Borkovec TD (2000) Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry* 31(2):73–86
- Dewey D, McDonald MK, Brown WJ, Boyd SJ, Bunnell BE, Schulberg D (2015) The impact of ecological momentary assessment on posttraumatic stress symptom trajectory. *Psychiatry Res* 230:300–303. <https://doi.org/10.1016/j.psychres.2015.09.009>
- Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D et al (2014) Focus article: report of the NIH task force on research standards for chronic low back pain
- Doblin R (1991) Pahnke's “Good Friday experiment”: a long-term follow-up and methodological critique. *J Transpers Psychol* 23:1–28
- Doering BK, Rief W, Petrie KJ (2014) Lessons to be learned from placebo arms in psychopharmacology trials. In: Benedetti F, Enck P, Frisaldi E, Schedlowski M (eds) *Placebo* Springer, Heidelberg, Berlin, pp 273–290. <https://doi.org/10.1007/978-3-662-44519-8>
- Doward LC, Gnanasakthy A, Baker MG (2010) Patient reported outcomes: looking beyond the label claim. *Health Qual Life Outcomes* 8:1–9. <https://doi.org/10.1186/1477-7525-8-89>
- Dragalin V (2011) An introduction to adaptive designs and adaptation in CNS trials. *Euro Neuropsychopharm* 21:153–158. <https://doi.org/10.1016/j.euroneuro.2010.09.004>
- Dupuis D (2021) Can psychedelics really change the world? Toward psychedelic technologies. MIND Foundation. <https://mind-foundation.org/psychedelic-technologies/>
- Dworkin RH, Turk DC, McDermott MP et al (2009) Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. *Pain* 146:238–244. <https://doi.org/10.1016/j.pain.2009.08.019>
- Dworkin RH, Turk DC, Peirce-Sandner S, et al. (2010) Research design considerations for confirmatory chronic pain clinical trials: IMMPACT recommendations. *Pain* 149:177–193. <https://doi.org/10.1016/j.pain.2010.02.018>
- Dworkin RH, Turk DC, Wyrwich KW et al (2008) Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 9:105–121. <https://doi.org/10.1016/j.jpain.2007.09.005>
- Earleywine M, Ueno LF, Mian MN, Altman BR (2021) Cannabis-induced oceanic boundlessness. *J Psychopharmacol*. <https://doi.org/10.1177/02698811211997099>
- Eaves ER, Sherman KJ, Ritenbaugh C, Hsu C, Nichter M, Turner JA, Cherkin DC (2015) A qualitative study of changes in expectations over time among patients with chronic low back pain seeking four CAM therapies. *BMC Complement Altern Med* 15:1–10. <https://doi.org/10.1186/s12906-015-0531-9>
- Eisner B (1997) Set, setting, and matrix. *J Psychoactive Drugs* 29:213–216. <https://doi.org/10.1080/02791072.1997.10400190>
- Enck P, Bingel U, Schedlowski M et al (2013) The placebo response in medicine: minimize, maximize or personalize? *Nat Rev Drug Discov* 12:191–204. <https://doi.org/10.1038/nrd3923>
- Enck P, Zipfel S (2019) Placebo effects in psychotherapy: a framework. *Front Psychiatry* 10:456. <https://doi.org/10.3389/fpsy.2019.00456>
- Erpelding N, Evans K, Lanier RK et al. (2020) Placebo response reduction and accurate pain reporting training reduces placebo responses in a clinical trial on chronic low back pain: results from a comparison to the literature. *Clin J Pain* 36:950–954. <https://doi.org/10.1097/AJP.0000000000000873>
- Evans K, Colloca L, Pecina M et al (2021) What can be done to control the placebo response in clinical trials? A narrative review. *Contemp Clin Trials* 107:106503. <https://doi.org/10.1016/j.cct.2021.106503>
- Fadiman J (2011) *The psychedelic explorer's guide: safe, therapeutic, and sacred journeys*. Simon and Schuster
- Ferrari R, Russell AS (2010) Effect of a symptom diary on symptom frequency and intensity in healthy subjects. *J Rheumatol* 37:2387–2389. <https://doi.org/10.3899/jrheum.100513>
- Fillingim RB, Price DD (2005) What is controlled for in placebo-controlled trials? In: *Mayo Clinic Proceedings* (Vol. 80, No. 9, pp. 1119–1121). Elsevier. <https://doi.org/10.4065/80.9.1119>
- Fisher S, Greenberg RP (1993) How sound is the double-blind design for evaluating psychotropic drugs? *J Nerv Ment Dis*. <https://doi.org/10.1097/00005053-199306000-00002>
- Fleming TR, Powers JH (2012) Biomarkers and surrogate endpoints in clinical trials. *Stat Med* 31:2973–2984. <https://doi.org/10.1002/sim.5403>
- Flowers EP, Freeman P, Gladwell VF (2018) Enhancing the acute psychological benefits of green exercise: An investigation of expectancy effects. *Psychology of Sport and Exercise* 39:213–221
- Freedland KE (2020) Purpose-guided trial design in health-related behavioral intervention research. *Health Psychol* 39:539–548. <https://doi.org/10.1037/hea0000867>
- Furukawa TA, Noma H, Caldwell DM et al (2014) Waiting list may be a nocebo condition in psychotherapy trials: a contribution

- from network meta-analysis. *Acta Psychiatr Scand* 130:181–192. <https://doi.org/10.1111/acps.12275>
- Gewandter JS, Dworkin RH, Turk DC et al (2020) Improving study conduct and data quality in clinical trials of chronic pain treatments: IMMPACT recommendations. *J Pain* 21:931–942. <https://doi.org/10.1016/j.jpain.2019.12.003>
- Glowacki DR, Wonnacott MD, Freire R et al (2020, April) Isness: using multi-person VR to design peak mystical type experiences comparable to psychedelics. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–14). <https://doi.org/10.1145/3313831.3376649>
- Gold SM, Enck P, Hasselmann H et al (2017) Control conditions for randomised trials of behavioural interventions in psychiatry: a decision framework. *Lancet Psychiatry* 4:725–732. [https://doi.org/10.1016/S2215-0366\(17\)30153-0](https://doi.org/10.1016/S2215-0366(17)30153-0)
- Griffiths RR, Johnson MW, Carducci MA et al (2016) Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: a randomized double-blind trial. *J Psychopharmacol* 30:1181–1197. <https://doi.org/10.1177/0269881116675513>
- Griffiths RR, Richards WA, McCann U, Jesse R (2006) Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. *Psychopharma* 187:268–283. <https://doi.org/10.1007/s00213-006-0457-5>
- Griffiths RR, Richards WA, Johnson MW, McCann UD, Jesse R (2008) Mystical-type experiences occasioned by psilocybin mediate the attribution of personal meaning and spiritual significance 14 months later. *J Psychopharmacol* 22(6):621–632
- Grob CS, Danforth AL, Chopra GS, Hagerty M, McKay CR, Halberstadt AL, Greer GR (2011) Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. *Arch Gen Psychiatry* 68:71–78
- Gukasyan N, Nayak SM (2021) Psychedelics, placebo effects, and set and setting: insights from common factors theory of psychotherapy. *Transcult Psych*. <https://doi.org/10.1001/archgenpsychiatry.2010.116>
- Haijen ECHM, Kaelen M, Roseman L et al (2018) Predicting responses to psychedelics: a prospective study. *Front Pharmacol* 9:897. <https://doi.org/10.3389/fphar.2018.00897>
- Hartogsohn (2021) American trip: set, setting, and the psychedelic experience in the twentieth century. The MIT Press, Cambridge
- Hartogsohn I (2018) The meaning-enhancing properties of psychedelics and their mediator role in psychedelic therapy, spirituality, and creativity. *Front Neurosci* 12:129. <https://doi.org/10.3389/fnins.2018.00129>
- Hartogsohn I (2016) Set and setting, psychedelics and the placebo response: an extra-pharmacological perspective on psychopharmacology. *J Psychopharmacol* 30:1259–1267. <https://doi.org/10.1177/0269881116677852>
- Hengartner MP (2020) Is there a genuine placebo effect in acute depression treatments? A reassessment of regression to the mean and spontaneous remission. *BMJ Evid Med* 25:46–48. <https://doi.org/10.1136/bmjebm-2019-111161>
- Heifets BD, Williams NR, Bentzley BS, Schatzberg AF (2019) Rigorous trial design is essential to understand the role of opioid receptors in ketamine's antidepressant effect. *JAMA Psych* 76:657–658. <https://doi.org/10.1001/jamapsychiatry.2019.0766>
- Hieronimus F, Lisinski A, Nilsson S, Eriksson E (2018) Efficacy of selective serotonin reuptake inhibitors in the absence of side effects: a mega-analysis of citalopram and paroxetine in adult depression. *Mol Psychiatry* 23:1731–1736. <https://doi.org/10.1038/mp.2017.147>
- Higgins JP, Altman DG, Gøtzsche PC et al (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343. <https://doi.org/10.1136/bmj.d5928>
- Holbein MB, Berglund JP, Weatherwax K (2015) Access to investigational drugs: FDA expanded access programs or “right to try” legislation? *Clin Transl Sci* 8:526–532. <https://doi.org/10.1111/cts.12255>
- Horton DM, Morrison B, Schmidt J (2021) Systematized review of psychotherapeutic components of psilocybin-assisted psychotherapy. *Amer J Psychother*. <https://doi.org/10.1176/appi.psychotherapy.20200055>
- Horvath AO, Del Re AC, Flückiger C, Symonds D (2011) Alliance in individual psychotherapy. *Psychother* 48:9–16. <https://doi.org/10.1037/a0022186>
- Hróbjartsson A, Gøtzsche PC (2010) Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev*. <https://doi.org/10.1002/14651858.CD003974.pub3>
- Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaut P, Brorson S (2012) Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 344:e1119. <https://doi.org/10.1136/bmj.e1119>
- Ivanova A, Zhang Z, Thompson L, Yang Y, Kotz RM, Fang X (2016) Can sequential parallel comparison design and two-way enriched design be useful in medical device clinical trials? *J Biopharm Stat* 26:167–177. <https://doi.org/10.1080/10543406.2015.1092028>
- Johnson MW, Griffiths RR (2017) Potential therapeutic effects of psilocybin. *Neurotherapeutics* 14(3):734–740
- Johnson MW, Garcia-Romeu A, Cosimano MP, Griffiths RR (2014) Pilot study of the 5-HT<sub>2A</sub>R agonist psilocybin in the treatment of tobacco addiction. *J Psychopharmacol* 28(11):983–992
- Jones BDM, Razza LB, Weissman CR et al (2021) Magnitude of the placebo response across treatment modalities used for treatment-resistant depression in adults: a systematic review and meta-analysis. *JAMA Netw Open* 4:e2125531. <https://doi.org/10.1001/jamanetworkopen.2021.25531>
- Juul S, Gluud C, Simonsen S, Frandsen FW, Kirsch I, Jakobsen JC (2020) Blinding in randomised clinical trials of psychological interventions: a retrospective study of published trial reports. *BMJ Evid Based Med* <https://doi.org/10.1136/bmjebm-2020-111407>
- Kaertner LS, Steinborn MB, Kettner H et al (2021) Positive expectations predict improved mental-health outcomes linked to psychedelic microdosing. *Sci Rep* 11:1–11. <https://doi.org/10.1038/s41598-021-81446-7>
- Kaptschuk TJ, Hemond CC, Miller FG (2020) Placebos in chronic pain: evidence, theory, ethics, and use in clinical practice. *BMJ* 370. <https://doi.org/10.1136/bmj.m1668>
- Katz N (2021) Design and conduct of confirmatory chronic pain clinical trials. *Pain Reports* 6. <https://doi.org/10.1097/PR9.0000000000000854>
- Kobak KA, Kane JM, Thase ME et al (2007) Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol* 27:1–5. <https://doi.org/10.1097/JCP.0b013e31802eb4b7>
- Kogo Y (2002) Aum Shinrikyo and spiritual emergency. *J Humanist Psychol* 42:82–101. <https://doi.org/10.1080/03060497.2003.11086195>
- Kolahi J, Bang H, Park J (2009) Towards a proposal for assessment of blinding success in clinical trials: up to date review. *Comm Dent Oral Epidem* 37:477–484. <https://doi.org/10.1111/j.1600-0528.2009.00494.x>
- Kube T, Rief W (2017) Are placebo and drug-specific effects additive? Questioning basic assumptions of double-blinded randomized clinical trials and presenting novel study designs. *Drug Discov Today* 22:729–735. <https://doi.org/10.1016/j.drudis.2016.11.022>
- Kuypers KP, Ng L, Erritzoe D et al (2019) Microdosing psychedelics: more questions than answers? An overview and suggestions for

- future research. *J Psychopharmacol* 33:1039–1057. <https://doi.org/10.1177/0269881119857204>
- Laferton JAC, Kube T, Salzmann S, Auer CJ, Shedden-Mora MC (2017) Patients' expectations regarding medical treatment: a critical review of concepts and their assessment. *Front Psychol* 8:233. <https://doi.org/10.3389/fpsyg.2017.00233>
- Leake GJ, King AS (1977) Effect of counsellor expectation on alcoholic recovery. *Alcohol, Health, and Research Work* 11(16)
- Leary T, Litwin GH, Metzner R (1963) Reactions to psilocybin administered in a supportive environment. *J Nerv Ment Dis* 137:561–573. <https://doi.org/10.1097/00005053-196312000-00007>
- Ledford JA (2019) Sex, drugs, and mind control: Tom O'Neill, with Dan Pipenbring: Chaos: Charles Manson, the CIA, and the Secret History of the Sixties Little, Brown and Company, New York
- Leger RF, Unterwald EM (2021) Assessing the effects of methodological differences on outcomes in the use of psychedelics in the treatment of anxiety and depressive disorders: a systematic review and meta-analysis. *J Psychopharmacol*. <https://doi.org/10.1177/02698811211044688>
- Lynn SJ, Evans J (2017) Hypnotic suggestion produces mystical-type experiences in the laboratory: a demonstration proof. *Psych Conscious: Theory Res Prac* 4:23–37. <https://doi.org/10.1037/cns0000105>
- MacLean KA, Leoutsakos JM, Johnson MW, Griffiths RR (2012) Factor analysis of the mystical experience questionnaire: a study of experiences occasioned by the hallucinogen psilocybin. *J Sci Study Relig* 51:721–737. <https://doi.org/10.1111/j.1468-5906.2012.01685.x>
- Majić T, Schmidt TT, Gallinat J (2015) Peak experiences and the afterglow phenomenon: when and how do therapeutic effects of hallucinogens depend on psychedelic experiences? *J Psychopharmacol* 29:241–253. <https://doi.org/10.1177/0269881114568040>
- Marks M, Cohen IG (2021) Psychedelic therapy: a roadmap for wider acceptance and utilization. *Nat Med*. <https://doi.org/10.1038/s41591-021-01530-3>
- Marton T, Barnes DE, Wallace A, Woolley JD (2019) Concurrent use of buprenorphine, methadone, or naltrexone does not inhibit ketamine's antidepressant activity. *Biological Psych* 85:e75–e76. <https://doi.org/10.1016/j.biopsych.2019.02.008>
- Mataix-Cols D, Andersson E (2021) Ten practical recommendations for improving blinding integrity and reporting in psychotherapy trials. *JAMA Psych*. <https://doi.org/10.1001/jamapsychiatry.2021.1419>
- Mathieu E, Herbert RD, McGeechan K, Herbert JJ, Barratt AL (2014) A theoretical analysis showed that blinding cannot eliminate potential for bias associated with beliefs about allocation in randomized clinical trials. *J Clin Epidemiol* 67:667–671. <https://doi.org/10.1016/j.jclinepi.2014.02.001>
- McCarney R, Warner J, Iliffe S et al (2007) The Hawthorne effect: a randomised, controlled trial. *BMC Med Res Methodol* 7:1–8. <https://doi.org/10.1186/1471-2288-7-30>
- Mercieca-Bebber R, King MT, Calvert MJ et al (2018) The importance of patient-reported outcomes in clinical trials and strategies for future optimization. *Patient Relat Outcome Meas* 9:353. <https://doi.org/10.1186/10.2147/PROM.S156279>
- Miller FG, Wendler D, Swartzman LC (2005) Deception in research on the placebo effect. *PLoS Med* 2:e262. <https://doi.org/10.1371/journal.pmed.0020262>
- Mithoefer MC, Feduccia AA, Jerome L et al (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. *Psychopharm* 236:2735–2745. <https://doi.org/10.1007/s00213-019-05249-5>
- Moseley JB, O'Malley K, Petersen NJ et al (2002) A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *NE J Med* 347:81–88. <https://doi.org/10.1056/NEJMoa013259>
- Mundt JM, Roditi D, Robinson ME (2017) A comparison of deceptive and non-deceptive placebo analgesia: efficacy and ethical consequences. *Ann Behav Med* 51:307–315. <https://doi.org/10.1007/s12160-016-9854-0>
- Murphy-Beiner A, Soar K (2020) Ayahuasca's 'afterglow': improved mindfulness and cognitive flexibility in ayahuasca drinkers. *Psychopharm* 237:1161–1169. <https://doi.org/10.1007/s00213-019-05445-3>
- Muthukumaraswamy S, Forsyth A, Lumley T (2021) Blinding and expectancy confounds in psychedelic randomised controlled trials. *Expert Rev Clin Pharmacol*. <https://doi.org/10.1080/17512433.2021.1933434>
- Oehen P, Traber R, Widmer V, Schnyder U (2013) A randomized, controlled pilot study of MDMA ( $\pm$ 3, 4-methylenedioxyamphetamine)-assisted psychotherapy for treatment of resistant, chronic Post-Traumatic Stress Disorder (PTSD). *J Psychopharm* 27:40–52. <https://doi.org/10.1177/0269881112464827>
- Olson DE (2020) The subjective effects of psychedelics may not be necessary for their enduring therapeutic effects. *ACS Pharm Trans Sci* 4:563–567. <https://doi.org/10.1021/acspsci.0c00192>
- Olson JA, Suissa-Rochelleau L, Lifshitz M et al (2020) Tripping on nothing: placebo psychedelics and contextual factors. *Psychopharm* 237:1371–1382. <https://doi.org/10.1007/s00213-020-05464-5>
- Pace BA, Devenot N (2021) Right-wing psychedelia: case studies in cultural plasticity and political pluripotency. *Front Psychol*
- Pahnke WN (1963) Drugs and mysticism: an analysis of the relationship between psychedelic drugs and the mystical consciousness: a thesis. Dissertation. Harvard University
- Pallmann P, Bedding AW, Choodari-Oskooei B et al (2018) Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med* 16:1–15. <https://doi.org/10.1186/s12916-018-1017-7>
- Patterson B, Boyle MH, Kivlenieks M, Van Ameringen M (2016) The use of waitlists as control conditions in anxiety disorders research. *J Psychiatr Res* 83:112–120. <https://doi.org/10.1016/j.jpsychires.2016.08.015>
- Peciña M, Bohnert AS, Sikora M, Avery ET, Langenecker SA, Mickey BJ, Zubieta JK (2015) Association between placebo-activated neural systems and antidepressant responses: neurochemistry of placebo effects in major depression. *JAMA Psych* 72:1087–1094. <https://doi.org/10.1001/jamapsychiatry.2015.1335>
- Peciña M, Zubieta JK (2013) Molecular mechanisms of placebo responses in humans. *Mol Psychiatry* 20:416–423. <https://doi.org/10.1038/mp.2014.164>
- Pedersen ER, Kaysen DL, Lindgren KP, Blayney J, Simpson TL (2014) Impact of daily assessments on distress and PTSD symptoms in trauma-exposed women. *J Interpers Violence* 29:824–845. <https://doi.org/10.1177/0886260513505705>
- Planès S, Villier C, Mallaret M (2016) The nocebo effect of drugs. *Pharmacol Res Perspect* 4:e00208. <https://doi.org/10.1002/prp2.208>
- Pollan M (2018) How to change your mind: what the new science of psychedelics teaches us about consciousness, dying, and addiction. Penguin, New York
- Prideaux, E. (2021). The worldview-changing drugs poised to go mainstream. BBCpage. <https://www.bbc.com/future/article/20210906-what-if-everyone-took-psychedelics>
- Puente I (2014) Holotropic breathwork can occasion mystical experiences in the context of a daylong workshop. *J Transper Res* 6:40–50
- Rabkin JG, Markowitz JS, Stewart JW et al (1986) How blind is blind? Assessment of patient and doctor medication guesses in a placebo-controlled trial of imipramine and phenelzine. *Psychiatry Res* 19:75–86. [https://doi.org/10.1016/0165-1781\(86\)90094-6](https://doi.org/10.1016/0165-1781(86)90094-6)

- Reckweg J, Mason NL, van Leeuwen C et al (2021) A phase 1, dose-ranging study to assess safety and psychoactive effects of a vaporized 5-methoxy-N, N-dimethyltryptamine formulation (GH001) in healthy volunteers. *Front Pharmacol* 12:760671. <https://doi.org/10.3389/fphar.2021.760671>
- Reissig CJ, Carter LP, Johnson MW, Mintzer MZ, Klinedinst MA et al (2012) High doses of extromethorphan, an NMDA antagonist, produce effects similar to classic hallucinogens. *Psychopharmacology* 223(1):1–15
- Rohsenow DJ, Marlatt GA (1981) The balanced placebo design: methodological considerations. *Addict Behav* 6:107–122. [https://doi.org/10.1016/0306-4603\(81\)90003-4](https://doi.org/10.1016/0306-4603(81)90003-4)
- Rosenkranz MA, Dunne JD, Davidson RJ (2019) The next generation of mindfulness-based intervention research: what have we learned and where are we headed? *Curr Opin Psychol* 28:179–183. <https://doi.org/10.1016/j.copsyc.2018.12.022>
- Rosenthal D, Frank JD (1956) Psychotherapy and the placebo effect. *Psychol Bull* 53:294–302. <https://doi.org/10.1037/h0044068>
- Ross S, Bossis A, Guss J et al (2016) Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: a randomized controlled trial. *J Psychopharmacol* 30:1165–1180. <https://doi.org/10.1177/0269881116675512>
- Roulet TJ, Gill MJ, Stenger S, Gill DJ (2017) Reconsidering the value of covert research: the role of ambiguous consent in participant observation. *Organ Res Methods* 20:487–517. <https://doi.org/10.1177/1094428117698745>
- Russ SL, Elliott MS (2017) Antecedents of mystical experience and dread in intensive meditation. *Psych Conscious Theory Res Prac* 4:38–53. <https://doi.org/10.1037/cns0000119>
- Rutherford BR, Roose SP (2013) A model of placebo response in antidepressant clinical trials. *Amer J Psych* 170:723–733. <https://doi.org/10.1176/appi.ajp.2012.12040474>
- Rutter M (2007) Proceeding from observed correlation to causal inference: the use of natural experiments. *Perspect Psychol Sci* 2:377–395. <https://doi.org/10.1111/j.1745-6916.2007.00050.x>
- Sackett DL (2007) Commentary: measuring the success of blinding in RCTs: don't, must, can't or needn't? *Int J Epidemiol* 36:664–665. <https://doi.org/10.1093/ije/dym088>
- Schenberg EE (2021) Who is blind in psychedelic research? Letter to the editor regarding: blinding and expectancy confounds in psychedelic randomized controlled trials. *Exp Rev Clin Pharm*. <https://doi.org/10.1080/17512433.2021.1951473>
- Scott AJ, Sharpe L, Colagiuri B (2022) A systematic review and meta-analysis of the success of blinding in antidepressant RCTs. *Psychiatry Res* 114297. <https://doi.org/10.1016/j.psychres.2021.114297>
- Scott AJ, Sharpe L, Quinn V et al (2021) Association of single-blind placebo run-in periods with the placebo response in randomized clinical trials of antidepressants: a systematic review and meta-analysis. *JAMA Psychiatry*. <https://doi.org/10.1001/jamapsychiatry.2021.3204>
- Sedgwick P, Greenwood N (2015) Understanding the Hawthorne effect. *Brit Med J* 351
- Sessa B (2014) Why psychiatry needs psychedelics and psychedelics need psychiatry. *J Psychoactive Drugs* 46:57–62. <https://doi.org/10.1080/02791072.2014.877322>
- Sherman KJ, Cherkin DC, Ichikawa L et al (2010) Treatment expectations and preferences as predictors of outcome of acupuncture for chronic back pain. *Spine* 35:1471–1477. <https://doi.org/10.1097/BRS.0b013e3181c2a8d3>
- Siegel AN, Meshkat S, Benitah K, Lipsitz O et al (2021) Registered clinical studies investigating psychedelic drugs for psychiatric disorders. *J Psychiatr Res*. <https://doi.org/10.1016/j.jpsychires.2021.05.019>
- Solowij N, Broyd S, Greenwood LM et al (2019) A randomised controlled trial of vaporised  $\Delta$  9-tetrahydrocannabinol and cannabidiol alone and in combination in frequent and infrequent cannabis users: acute intoxication effects. *Eur Arch Psychiatry Clin Neurosci* 269:17–35
- Smart RG, Storm T, Baker EF, Solursh L (1966) A controlled study of lysergide in the treatment of alcoholism. I. The effects on drinking behavior. *Q J Stud Alc* 27:469–482
- Smith SM, Dworkin RH, Turk DC et al (2017) The potential role of sensory testing, skin biopsy, and functional brain imaging as biomarkers in chronic pain clinical trials: IMMPACT considerations. *J Pain* 18:757–777. <https://doi.org/10.1016/j.jpain.2017.02.429>
- Smith SM, Dworkin RH, Turk DC et al (2020) Interpretation of chronic pain clinical trial outcomes: IMMPACT recommended considerations. *Pain* 161:2446–2461. <https://doi.org/10.1097/j.pain.0000000000001952>
- Smith WR, Sisti D (2021) Ethics and ego dissolution: the case of psilocybin. *J Med Ethics* 47:807–814. <https://doi.org/10.1136/medethics-2020-106070>
- Strimbu K, Tavel JA (2010) What are biomarkers? *Curr Opin HIV AIDS* 5:463. <https://doi.org/10.1097/COH.0b013e32833ed177>
- Szigeti B, Kartner L, Blemings A et al (2021) Self-blinding citizen science to explore psychedelic microdosing. *Elife* 10:e62878. <https://doi.org/10.7554/eLife.62878>
- Tambling RB (2012) A literature review of therapeutic expectancy effects. *Contemp Fam Ther* 34:402–415. <https://doi.org/10.1007/s10591-012-9201-y>
- Tamura RN, Huang X (2007) An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. *Clin Trials* 4:309–317. <https://doi.org/10.1177/1740774507081217>
- Thapar A, Rutter M (2019) Do natural experiments have an important future in the study of mental disorders? *Psychol Med* 49:1079–1088. <https://doi.org/10.1017/S0033291718003896>
- Timmermann C, Kettner H, Letheby C, Roseman L, Rosas F, Carhart-Harris R (2021) Psychedelics alter metaphysical beliefs. [10.31234/osf.io/f6sjk](https://doi.org/10.31234/osf.io/f6sjk)
- Turner JA, Deyo RA, Loeser JD, Von Korff M, Fordyce WE (1994) The importance of placebo effects in pain treatment and research. *J Amer Med Assoc* 271:1609–1614. <https://doi.org/10.1001/jama.1994.03510440069036>
- US Food and Drug Administration (2009) Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. Silver Spring, MD
- US Food and Drug Administration (2020) Guidance for industry and FDA staff. Qualification Process for Drug Development Tools, Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf>
- Uthaug MV, Mason NL, Toennes SW et al (2021) A placebo-controlled study of the effects of ayahuasca, set and setting on mental health of participants in ayahuasca group retreats. *Psychopharm*. <https://doi.org/10.1007/s00213-021-05817-8>
- van Elk M, Fejer G, Lempe P, Prochazkova L, Kuchar M, Hajkova K, Marshall J (2021) Effects of psilocybin microdosing on awe and aesthetic experiences: a preregistered field and lab-based study. *Psychopharm* 1–16. <https://doi.org/10.1007/s00213-021-05857-0>
- Vase L, Wartolowska K (2019) Pain, placebo, and test of treatment efficacy: a narrative review. *Brit J Anaes* 12:e254–e262. <https://doi.org/10.1016/j.bja.2019.01.040>
- Višlā A, Constantino MJ, Newkirk K, Ogródniczuk JS, Söchting I (2018) The relation between outcome expectation, therapeutic alliance and outcome among depressed patients in group cognitive-behavioral therapy. *Psychother Res* 28:446–456. <https://doi.org/10.1080/10503307.2016.1218089>

- Wampold BE, Frost ND, Yulish NE (2016) Placebo effects in psychotherapy: a flawed concept and a contorted history. *Psych Consc Theory Res Prac* 3:108–120. <https://doi.org/10.1037/cns0000045>
- Webb CA, DeRubeis RJ, Barber JP (2010) Therapist adherence/competence and treatment outcome: a meta-analytic review. *J Consult Clin Psychol* 78:200–211. <https://doi.org/10.1037/a0018912>
- Weiss B, Miller JD, Carter NT, Campbell WK (2021) Examining changes in personality following shamanic ceremonial use of ayahuasca. *Sci Rep* 11:1–15. <https://doi.org/10.1038/s41598-021-84746-0>
- Whitney CW, Von Korff M (1992) Regression to the mean in treated versus untreated chronic pain. *Pain* 50:281–285. [https://doi.org/10.1016/0304-3959\(92\)90032-7](https://doi.org/10.1016/0304-3959(92)90032-7)
- Williams NR, Heifets BD, Bentzley BS, Blasey C, Sudheimer KD, Hawkins J, Lyons DM, Schatzberg AF (2019) Attenuation of antidepressant and antisuicidal effects of ketamine by opioid receptor antagonism. *Mol Psychiatry* 24:1779–1786. <https://doi.org/10.1038/s41380-019-0503-4>
- Williams NR, Heifets BD, Blasey C et al (2018) Attenuation of antidepressant effects of ketamine by opioid receptor antagonism. *Am J Psychiatry* 175:1205–1215. <https://doi.org/10.1176/appi.ajp.2018.18020138>
- Williams MT, Labate BC (2020) Diversity, equity, and access in psychedelic medicine. *J Psychedelic Stud* 4:1–3. <https://doi.org/10.1556/2054.2019.032>
- Wilsey B, Deutsch R, Marcotte TD (2016) Maintenance of blinding in clinical trials and the implications for studying analgesia using cannabinoids. *Cannabis Cannabinoid Res* 1:139–148. <https://doi.org/10.1089/can.2016.0016>
- Wolfson PE, Andries J, Feduccia AA et al (2020) MDMA-assisted psychotherapy for treatment of anxiety and other psychological distress related to life-threatening illnesses: a randomized pilot study. *Sci Rep* 10:1–15. <https://doi.org/10.1038/s41598-020-75706-1>
- Yaden DB, Griffiths RR (2020) The subjective effects of psychedelics are necessary for their enduring therapeutic effects. *ACS Pharm Trans Sci* 4:568–572. <https://doi.org/10.1021/acspsci.0c00194>
- Yockey RA, Vidourek RA, King KA (2020) Trends in LSD use among US adults: 2015–2018. *Drug Alcohol Depend* 212:108071. <https://doi.org/10.1016/j.drugalcdep.2020.108071>
- Yoo SK, Hong S, Sohn N, O'Brien KM (2014) Working alliance as a mediator and moderator between expectations for counseling success and counseling outcome among Korean clients. *Asia Pac Educ Rev* 15:271–281. <https://doi.org/10.1007/s12564-014-9320-2>
- Yoon G, Petrakis IL, Krystal JH (2019) Association of combined naltrexone and ketamine with depressive symptoms in a case series of patients with depression and alcohol use disorder. *JAMA Psych* 76:337–338. <https://doi.org/10.1001/jamapsychiatry.2018.3990>
- Younger J, Gandhi V, Hubbard E, Mackey S (2012) Development of the Stanford Expectations of Treatment Scale (SETS): a tool for measuring patient outcome expectancy in clinical trials. *Clinical Trials* 9(6):767–776
- Zhipei ZHU, Zhang L, Jiang J et al (2014) Comparison of psychological placebo and waiting list control conditions in the assessment of cognitive behavioral therapy for the treatment of generalized anxiety disorder: a meta-analysis. *Shanghai Arch Psychiatry* 26(319). <https://doi.org/10.11919/j.issn.1002-0829.214173>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.