universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The Effectiveness of Psychedelic Breakthrough Therapies."

A Three-Level Meta-Analysis on the Effects of Psilocybin- and MDMA-Assisted Psychotherapy

verfasst von / submitted by

Franziska Kätzel, BSc

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree program code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / degree program as it appears on the student record sheet:	Masterstudium Psychologie UG2002
Betreut von / Supervisor:	Dr. Anton-Rupert Laireiter, Privatdoz.

Table of Content

List of Tables and Figures

List of Abbreviations

Abbreviation	Definition
AASE	Alcohol Abstinence Self-Efficacy Scale
AD	Anxiety disorder
APA	American Psychological Association
AR	Agreeableness rate
AUD	Alcohol use disorder
BDI	Beck Depression Inventory
CAPS	Clinician-administered PTSD Scale
CBCT	Cognitive-behavioral conjoint therapy
CBT	Cognitive behavioral therapy
CRAN	Comprehensive R Archive Network
DEA	Drug Enforcement Administration
DMN	Default mode network
DMT	N, N-dimethyltryptamine
FDA	Food and Drug Administration
HPPD	Hallucinogen-persisting perception disorder
ICTRP	International Clinical Trials Registry Platform
IES	Impact of Events Questionnaire
IRR	Interrater reliability
ISCD	Independent Scientific Committee on Drugs
LSD	Lysergic Acid Diethylamide
LTI	Life-threatening illness
LTFU	Long-term follow-up
MA	Meta-analysis
MAPS	Multidisciplinary Association for Psychedelic Science
MDD	Major depressive disorder
MDMA	3,4-methylenedioxymethamphetamine
PACS	Penn Alcohol Craving Scale
PDS	Posttraumatic Diagnostic Scale
PICO	Population Intervention Comparators Outcomes

PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analyses
PS	Primary study
PTBS	Posttraumatischer Belastungsstörung
PTSD	Post-traumatic stress disorder
QIDS	Quick Inventory of Depressive Symptomatology
REML	REstricted Maximum Likelihood
RCT	Randomized controlled trial
RoB	Risk of Bias
SSRI	Selective serotonin reuptake inhibitor
STAI	State-Trait Anxiety Inventory
SUD	Substance use disorder
TAU	Treatment as usual
TRD	Treatment-resistant depression
WHO	World Health Organization

Abstract

Background

Established pharmacotherapies for depression and post-traumatic stress disorder (PTSD) show limited efficacy, while preliminary results of psilocybin- and 3,4-methylenedioxy methamphetamine (MDMA)-assisted psychotherapy are promising. The US Food and Drug Administration (FDA) designated those recent treatment approaches as 'breakthrough therapies'.

Objective

This meta-analysis uses the novel three-level model (a meta-analytic approach to account for dependency in effect sizes by differentiation of between- and within-study effect size variation) to quantitively summarize the acute and persisting overall effects of psilocybin- and MDMA-assisted therapy studies.

Methods

In total, eight databases were searched for published and grey literature. Clinical trials (Randomized controlled trials [RCTs] and open-label designs) that involved the administration of psilocybin and MDMA to patients suffering from depression, anxiety, PTSD, and alcohol use disorder were considered eligible. For the primary endpoints, mean scores of instruments assessing the severity of symptoms of psychological disorders were extracted before and after treatment with psychedelic substances. A pre-post short-term and long-term summary effect was calculated using Hedges' g.

Results

16 primary studies with a total number of 463 participants were included in the meta-analysis. The short-term summary effect (1-6 weeks post treatment) was estimated as Hedges' $g = 1.684$ ($p < .001$) and the long-term summary effect (6-12 months post treatment) as Hedges' $g = 1.650$ ($p < .001$). The short-term summary effect was significantly moderated by the type of substance ($p = .028$) and type of disorder ($p = .002$), but not by the study design used ($p = .211$) or the estimated study quality ($p = .283$).

Conclusion

Psilocybin and MDMA produce significant and very large effects in treatment settings that persist over six to twelve months. Even though the interpretability of results is still limited due to studies with mostly small and selective samples, psilocybin and especially MDMA studies show great potential for prospective psychopharmacological solutions.

Keywords: three-level, meta-analysis, psychedelic-assisted psychotherapy, psilocybin, MDMA

Abstract (German)

Hintergrund

Gängige Psychopharmaka zur Behandlung von Depression und posttraumatischer Belastungsstörung (PTBS) weisen nur eine begrenzte Wirksamkeit auf. Jüngste Studien der Psilocybin- und 3,4-methylenedioxymethamphetamine (MDMA)- gestützten Psychotherapie zeigen dagegen vielversprechende Ergebnisse. Die US Food and Drug Administration (FDA) hat diese neuen Behandlungsansätze zu ‚Breakthrough Therapies' erklärt.

Zielsetzung

Diese Meta-analyse wendet das Three-level Modell an (ein metaanalytischer Ansatz zur Berücksichtigung der Abhängigkeiten von Effektstärken), um die akuten und nachhaltigen Gesamteffekte von Psilocybin- und MDMA- unterstützten Therapiestudien zu erfassen.

Methoden

Es wurden insgesamt acht Datenbanken nach veröffentlichter und grauer Literatur durchsucht. Klinische Studien (RCTs und Open-Label Studien), die Psilocybin und MDMA zur Behandlung von Depressionen, Angststörungen, PTBS und Alkoholabhängigkeit anwendeten, wurden inkludiert. Symptomreduktion der psychischen Störungen, erhoben mit klinisch-psychologischen Instrumenten vor und nach der Behandlung mit psychedelischen Substanzen, dienten als abhängige Variablen. Ein kurzfristiger und langfristiger prä-post Gesamteffekt wurde mit Hilfe von Hedges' g berechnet.

Ergebnisse

16 Primärstudien mit insgesamt 463 Teilnehmer*innen wurden in die Meta-analyse integriert. Der kurzfristige Gesamteffekt (1-6 Wochen nach Behandlung) wurde mit Hedges' g = 1.684 (p < .001) und der langfristige Gesamteffekt (6-12 Monate nach Behandlung) mit Hedges' g = 1.650 (p < .001) geschätzt. Der kurzfristige Gesamteffekt wurde signifikant durch die Art der Substanz (p = .028) und der Störung (p = .002) moderiert, nicht aber durch das verwendete Studiendesign (p = .211) oder die geschätzte Studienqualität (p = .283).

Schlussfolgerung

Psilocybin und MDMA haben in Behandlungssettings signifikante und große Wirkungen, die über sechs bis zwölf Monate anhalten. Die Interpretierbarkeit der Ergebnisse ist aufgrund von Studien mit kleinen und selektiven Stichproben limitiert. Dennoch haben Psilocybin- und insbesondere MDMA-Studien ein großes psychopharmakologisches Potential.

Keywords: three-level, meta-analysis, psychedelic-assisted psychotherapy, psilocybin, MDMA

Introduction

Recent crises have caused an acute increase in mental health conditions globally. Due to the COVID-19 pandemic, the prevalence of anxiety and depression rose by 25% worldwide (World Health Organization [WHO], 2022). Other crises, like the war in Ukraine, cause trauma for millions of people (Euronews, 2022).

Although established pharmacotherapies for depression and post-traumatic stress disorder (PTSD) are on the market, treatments show limited efficacy (Kolovos et al., 2017). In the meta-analysis of Kolovos et al. (2017), pharmacological treatment as usual (TAU) for major depression was only effective for one-third of the patients. Likewise, the large-scale out-patient study STAR*D showed that only one-third of the patients achieved full symptom relief when treated with the selective serotonin reuptake inhibitor (SSRI) citalopram (Trivedi et al., 2006). Medical treatments for PTSD display similar shortcomings: although benzodiazepines are commonly used as a treatment for PTSD, their risks seem to outweigh the potential benefits (Guina et al., 2015). Though evidence supports the efficacy of common SSRIs (fluoxetine, paroxetine, and sertraline) for improving PTSD symptoms, the effects are limited (Jonas et al., 2013).

A new approach to treating mental disorders has attracted a lot of attention in the past decade: psychedelic substances. According to Albert Hofmann (the discoverer of the psychedelic effects of lysergic acid diethylamide [LSD]), the longing for spirituality and holistic healing is an understandable reaction to a materialistic and industrialized society that has lost its connection to nature. Albert Hofmann believed that psychedelics would provide civilization with a more holistic therapeutic solution (Pollan, 2019).

Substances

This present meta-analysis is focusing on the two psychoactive substances that are, to date, the broadest researched in the context of psychotherapy. These substances are the US Food and Drug Administration (FDA) designated ‘breakthrough therapies’ psilocybin and 3,4-methylenedioxymethamphetamine (MDMA). This FDA designation means that preliminary results are promising and supported throughout the clinical development process (Doblin et al., 2019). The following chapter will briefly overview the neurobiological properties of these drugs and their mechanisms of action on a cognitive, emotional, and behavioral level. Even though MDMA is not a classic psychedelic substance (see chapter MDMA), it will be referred to as a ‘psychedelic’ in the following course of this thesis, to ease the reading experience.

Psilocybin

Psilocybin is a hallucinogenic chemical that occurs naturally in almost 150 species of fungi naturally growing in Europe, South America, Mexico, and the United States (Rätsch, 2018). Mushrooms that contain psilocybin are commonly known as magic mushrooms. Psilocybin is classified as a psychedelic drug alongside LSD, 3,4,5-trimethoxyphenethylamine (mescaline), and N, N-dimethyltryptamine (DMT). These substances induce temporary alterations in mood, perception, and cognition through the activation of serotonin receptors in the brain (Rucker et al., 2020). Findings show that the 5-HT2A receptor is responsible for the mediation of hallucinogenic effects in human subjects (Halberstadt, 2015; Nichols, 2016). Psilocybin acts as a serotonin agonist in frontal and limbic areas, increasing glutamatergic transmission and neuroplasticity. These neurochemical effects are related to changes in self-perception, increased introspection, and positive mood. Furthermore, the psychoactive effects of psilocybin can be associated with increases in the personality trait 'openness', empathy, and improvements in emotional processing (Dos Santos & Hallak, 2020).

Recent studies have examined the effects of psilocybin on the default mode network (DMN). The DMN is a superordinate intrinsic brain network associated with cognitive functions like introspection and self-referential thinking (Andrews-Hanna et al., 2014). The DMN is often overactive in major depression, causing excessive rumination and rigid patterns of thinking (Hamilton et al., 2011). Recent findings show that the DMN became more functionally interconnected and flexible after psilocybin treatment, implying that psilocybin may increase global brain network integration and decrease depressive symptoms (Daws et al., 2022).

MDMA

MDMA is a psychoactive, synthetically obtained phenethylamine derivate with structural and pharmacological similarities to both amphetamines and mescaline. (Papaseit et al., 2018). In a recreational context, MDMA is commonly referred to as 'ecstasy', even though ecstasy pills are oftentimes impurified with other substances. The intake of MDMA is associated with a reduction of amygdala activity in response to negative emotional stimuli and increased connectivity between the amygdala and hippocampus (Carhart-Harris et al., 2015). Aside from the release of serotonin, dopamine, and norepinephrine, MDMA especially elevates levels of the neurohormone oxytocin (Nardou et al., 2019; Wolfson et al., 2020). This so-called 'empathogen' produces an emotional state with a heightened mood, reduced fear, intensified sensory perception, and increased levels of both empathy and prosocial behavior (Bedi et al., 2010; Liechti et al., 2000).

Recent results show that a single dose of MDMA reopens an oxytocin-dependent critical period of social reward learning that typically closes after adolescence (Nardou et al., 2019). In states of trauma, the closure of critical periods inhibits adaption of the brain, even when optimal conditions are restored. These findings suggest that MDMA may help the processing and release of intractable, fear-related memories (Nardou et al. 2019). It is discussed that MDMA, in combination with psychotherapy, may produce a 'window of tolerance,' in which participants can revisit and process traumatic memories without the symptoms of hyperarousal, dissociation, or overwhelming anxiety (Mitchell et al., 2021; Mithoefer et al., 2017).

The History of Psychedelic Science

"Psychedelics may be the oldest class of psychopharmacological agents known to man." (Nichols, 2016, p. 268)

Psychedelic science has a long history. The use of psilocybin, which is classified as a 'classic psychedelic' through its 5-HT2A receptor action, has deep roots in pre-Columbian Mesoamerican cultures. Ancient civilizations in Central America, including the Olmec, Zapotec, Maya, and Aztecs, regularly used various psychoactive plants for healing rituals and religious ceremonies (Carod-Artal, 2015). Aztec shamans, for example, practiced healing by inducing altered states of consciousness with the hallucinogenic effects of psilocybin mushrooms (Ott & Bigwood, 1978). The following chapter will briefly summarize the history of modern psychedelic research from its first wave in the 1930s until its so-called 'renaissance', which we are now in.

The First Wave of Psychedelic Science (1938 – 1970)

The discovery of LSD in 1938 by the Swiss chemist Albert Hofmann is commonly dated as the beginning of modern psychedelic research (Doblin et al., 2019). This event has a special significance in neuroscience because research on serotonin in brain functionality was catalyzed by the discovery of LSD (Nichols, 2016). Albert Hofmann also identified the active component psilocybin in magic mushrooms in 1958 (Hofmann et al., 1958). In the following years, psychoactive substances were researched extensively in a psychiatric context. By the end of the 1960s, hundreds of papers examined the use of mescaline, psilocybin, and (most frequently) LSD for depressive, anxious, obsessive, and addictive disorders (Rucker et al., 2018).

MDMA was first synthesized in 1912 by Merck, a pharmaceutical company, and in the 1970s was rediscovered by the chemist A Shulgin (Benzenhöfer & Passie, 2006). In contrast to early research of the 'classic psychedelics', only a few early studies were conducted on the potential effects of MDMA in clinical settings (Greer & Tolbert, 1998).

While recreational use of psychedelic substances increased throughout the 1960s, it reached its peak through the Vietnam-war protests (1965-1973). Psychedelic drugs were associated with cultural and political rebellion and therefore targeted by media and political campaigns (Rucker et al., 2018). Unethical and covert use, the emergence of the hallucinogen-persisting perception disorder (HPPD), and a general hardening of social-political attitudes towards drug use, led to the declaration of LSD and psilocybin as Schedule I drugs in the Controlled Substances Act of 1970 in the United States (Rucker et al., 2018). In 1985, MDMA was additionally placed on the list of Schedule I drugs by the US Drug Enforcement Administration (DEA). This designation means that these substances have no currently accepted medical use and a high potential for abuse (National Institute on Drug Abuse, 2017). Consequently, psychedelic research experienced a 30-year hiatus.

The Renaissance of Psychedelic Science (2000 – now)

Since the turn of the millennium, clinical and academic interest in psychoactive substances has reemerged and grown into a steady renaissance. Simultaneously, the socio-political narrative has slowly changed into questioning the relative benefits and harms of the so-called ‘war on drugs’ (Rucker et al., 2018). The new era of controlled psychedelic research has focused on more careful experimental designs and more critical interpretation of outcomes (Carhart-Harris & Goodwin, 2017). Within the growing body of neuroscience on psychedelics, modern psychedelic research has focused especially on clinical trials with MDMA-assisted therapy for the treatment of PTSD and psilocybin clinical studies for depression and addiction (Doblin et al., 2019).

Psilocybin- and MDMA-Assisted Psychotherapy

Psychotherapy with psychoactive substances typically entails three types of sessions: preparatory, treatment with moderate to high doses of the psychoactive drug, and integration sessions (Reiff et al., 2020). In the dosing sessions, psilocybin-assisted therapy aims to create an overwhelming and transcendent experience, a ‘peak experience’, with rather high doses of the substance (Garcia-Romeu & Richards, 2018). Contrary to that, MDMA-assisted therapy intends to reach an optimal degree of emotional engagement, avoiding dissociation or overwhelming emotions, alongside exposure (Foa et al., 2007). This is described as the ‘optimal arousal zone’ or ‘window of tolerance’ (Mithoefer et al., 2010). The course of psychedelic-assisted therapy sessions in the context of clinical trials will be illustrated in the following chapter.

The Course of Therapy Sessions

The preparatory period usually employs three sessions, in which the patient's history is gathered and several questionnaires are completed. The therapist provides necessary information about the drug's possible effects, including adverse effects, and addresses the participant's questions and concerns. The main goal of these sessions is to build rapport and trust between the patient and therapist and to increase the patient's sense of comfort and safety. It is also an opportunity to model relevant attitudes during dosing sessions, such as patience and respect for one's boundaries. Personal or family history of psychotic or bipolar disorders typically leads to exclusion from the trial. (Davis et al., 2021; Carhart-Harris et al., 2016; Griffiths et al., 2016; Mithoefer et al., 2017)

Treatment sessions can vary from one to three substance-dosing sessions, usually lasting six to eight hours. A comfortable environment with fresh flowers, low lights, and fabric wall hangings is provided in most clinical settings. Patients are instructed to lie down and focus inwards with blindfolds and carefully chosen music. The intention is to allow each participant's experience to unfold spontaneously without a specific agenda. The sessions' scope includes exploring psychological, interpersonal, and spiritual aspects of life. The therapist's approach is primarily non-directive and supportive. In psilocybin sessions, the therapist's role is mainly monitoring. In MDMA sessions, therapists also focus on empathic listening and occasional guidance in cases of emotional blockages. Blood pressure, body temperature, and heart rate are constantly monitored by physical practitioners. (Davis et al., 2021; Carhart-Harris et al., 2016; Griffiths et al., 2016; Mitchell et al., 2021; Mithoefer et al., 2017; Ot'alora et al., 2018;).

Integration ordinarily takes place in sessions between each drug administration and usually three sessions after treatment. Cognitive, affective, and psychospiritual experiences are discussed. The focus lies on new thoughts and feelings that arose during the dosing sessions. These experiences can later provide a basis from which the participant can approach the emergence of painful emotions with a greater sense of strength and safety. The goal is to create a shift in consciousness that allows the patient to develop a new sense of mastery over rigid thinking, trauma, or other dysfunctional behaviors and emotions. (Griffiths et al., 2016; Mithoefer et al., 2017; Ot'alora et al., 2018).

The Importance of Set and Setting

The previously described structure of psychedelic-assisted therapy is designed to create an optimal drug experience, considering the importance of set and setting. The umbrella terms 'set' and 'setting' are contextual factors that can determine the experience of a drug session. 'Set' corresponds to biological factors, personality traits, and states that can influence the

quality of the experience (Hartogsohn, 2016). Careful selection and preparation as well as discussion of expectation and intention can shape the effects of psychoactive substances and minimize adverse reactions in the participants (Hartogsohn, 2017). ‘Setting' refers to the physical, social, and cultural context factors of the experience (Hartogsohn, 2016). A safe and comfortable environment that encourages relaxation, openness, and trust can facilitate a pleasant experience (Hartogsohn, 2017).

Safety and Tolerability

The safety and tolerability of classic psychedelics have been reviewed more extensively than the safety and tolerability of MDMA. This is presumably due to the inherent hallucinogenic effects of classic psychedelics that are hard to control externally. Therefore, the following chapter will briefly review the safety and tolerability of psilocybin use.

According to Nutt (2021), it is nearly impossible to die on an overdose of psilocybin. Multiple studies and reviews report that psilocybin does not cause addiction (Gable, 1993; Halberstadt, 2015; Johansen & Krebs, 2015; Johnson et al., 2008; Nutt, 2021). A large population study of 130,000 adults in the United States claims that psychedelics neither harm body organs nor cause damage to the brain, and that serious adverse events are very rare (Johansen & Krebs, 2015). A pooled analysis with raw data from eight controlled trials using psilocybin in healthy subjects showed that psilocybin did not cause long-term perceptual, cognitive, or neurological side effects. Although the substance-induced profound changes in perception, most subjects reported an enriching and non-threatening experience (Studerus et al., 2010).

Compared to other drugs, psilocybin is reviewed to be the least harmful. Members of the Independent Scientific Committee on Drugs (ISCD) ranked 20 different drugs in the categories of self-harm and other-harm. While alcohol reached the highest ranking (followed by heroin and crack cocaine), psilocybin was rated as the least harmful drug at all (Nutt et al., 2010). An extensive review of addiction rates of these 20 substances also found that heroin has the greatest risk of dependence and acute lethality, while psilocybin has the least (Gable, 1993).

Nonetheless, the administration of psychedelics bears certain risks that should not be neglected: on rare occasions, high doses of psilocybin can induce acute adverse reactions, like strong dysphoria or anxiety (Studerus et al., 2010). According to Johnson et al. (2008), the most likely risk is overwhelming distress during drug action, which is commonly known as a ‘bad trip’. In very rare cases, the consumption of a psychedelic substance can cause HPPD. This syndrome is expressed by prolonged or reoccurring hallucinogen effects, commonly known as ‘flashbacks’, months or years after hallucinogen use (Orsolini et al., 2017). Current knowledge

of HPPD is limited. It is reported mostly after illicit LSD use, and less commonly in treatment settings or with the use of other types of hallucinogens, like psilocybin (Halpern & Pope, 2003).

In sum, potential harm occurs most often in unprepared, unsupervised users with a predisposition to psychotic disorders (Johnson et al., 2018). Safeguards against these risks are the exclusion of patients with a history of psychotic disorders, establishment of trust and rapport between patient and therapist, a safe physical setting, and control for the relatively rare HPPD in a follow-up contact. Persisting adverse reactions are unlikely when research is conducted along these guidelines (Johnson et al., 2008; Johnson et al., 2018; Studerus et al., 2010).

The Potential of Psilocybin- and MDMA-Assisted Therapy

What makes psychedelic-assisted therapy so special? One possible gain is the amplified patient-therapist-alliance. According to a task force of the American Psychological Association (APA) reviewing 16 meta-analyses, the relationship between patient and therapist plays an immense role in the effectiveness of psychotherapy (DeAngelis, 2019). As stated in the manual for MDMA-assisted psychotherapy, MDMA can enhance the therapeutic alliance due to increased openness and trust and, therefore, catalyze therapeutic processing (Mithoefer et al., 2017). In the following chapter, the potential advantages of psilocybin- and MDMA-assisted therapy over conventional treatment will be summarized.

Persisting Effects

Another possible advantage of psychedelic-assisted therapy is that symptom reduction is relatively sustainable. A systematic review of 34 trials using classic psychedelics (most frequently, administered psilocybin) demonstrates that changes in depression, anxiety, substance misuse, and well-being endure over a long period – until up to six months after the immediate drug effects have worn off. Peak experiences and emotional breakthroughs were associated with long-term effects (Aday et al., 2020). Similarly, a longitudinal pooled analysis of six MDMA trials shows that symptom reduction persisted for at least 12 months post-treatment (Jerome et al., 2020).

Efficacy through Peak Experiences

The term ‘peak experience’ was shaped by Maslow (1959), who described this phenomenon as a moment of the highest happiness and fulfillment that can occur in different forms of everyday life, namely the mystic, oceanic, or natural experience, the aesthetic perception, the creative moment, the therapeutic or intellectual insight, the orgasmic experience, and certain forms of athletic fulfillment.

According to Majić et al. (2015), diverse approaches in psychotherapy suggest that peak experiences can differentially mediate the effects of psychotherapy. One of the unique

properties of psychedelic substances lies in creating altered states of consciousness that resemble such peak experiences. Already in ancient rituals, the use of psychedelics showed effects through the enhancement of social interactions and pro-community attitudes, which was understood as part of holistic healing (Majić et al., 2015).

In line with these findings, the 'psychedelic treatment model' suggests that the quality of the psychedelic experience functions as a key mediator of sustainable symptom reduction (Roseman et al., 2018). In the study of Roseman et al. (2018), it was demonstrated that features of the mystical peak experience after the administration of psilocybin predicted long-term positive outcomes in patients with treatment-resistant depression (TRD).

The Current State of Research

With the pioneering work of the Multidisciplinary Association for Psychedelic Science (MAPS) and other independent research organizations, psychedelic research has increased drastically in the past decade. Positive media coverage and new economic interests have accelerated this hype (Petranker et al., 2020). Since the FDA's 'breakthrough therapy' designation, clinical trials with psilocybin and MDMA picked up speed running through trial phases for treatment admission (Doblin et al, 2019).

Trial Phases

To get FDA approval, a new drug must pass the three phases of clinical trials. 'Phase 1' clinical trials focus on the safety and tolerability of the substance, usually performed on a small group of healthy volunteers or patients with life-threatening illnesses. In 'phase 2', researchers concentrate on the most effective dosages for larger numbers of participants with treatment-relevant disorders. 'Phase 3' trials try to replicate the results of the earlier trials in larger patient groups, involving several hundred participants in many study locations. This is the final step before seeking FDA approval. (U.S. Food and Drug Administration, 2018)

Existing Meta-Analyses

A meta-analysis is a quantitative method to systematically summarize existing studies in a certain research field (Borenstein et al., 2021). Several meta-analyses and systematic reviews about the effects of psychedelic drugs have been published, integrating primary studies on the effects of classic psychedelics (Andersen et al., 2021; Dos Santos et al., 2016; Galvão-Coelho et al., 2021; Goldberg et al., 2020; Romeo et al., 2020) and MDMA (Amoroso & Workman, 2016; Bahji et al., 2020; Hoskins et al., 2021; Illingworth et al., 2021; Tedesco et al., 2021). One possible limitation of these existing meta-analyses could be the risk of overlapping information. When including multiple outcome variables for the same sample of

participants, there is a risk of unaccounted dependency in effect sizes (Van den Noortgate et al., 2015).

This issue was addressed by the unpublished meta-analysis of Leudesdorff (2021) as part of the same master's seminar supervised by Univ.-Prof. Dr. Phil. Anton-Rupert Laireiter. Leudesdorff conducted a meta-analysis on the effects of psychedelics- and MDMA-assisted therapy in clinical settings, including various substances (ayahuasca, ibogaine, psilocybin, LSD, and MDMA). To account for potential dependency in the different outcomes, Leudesdorff imputed an average correlation of $r = .50$.

To my knowledge, no meta-analysis in the field of psychedelic science has used the method of multilevel meta-analysis before. This statistical model is known to be a new and strong methodological approach to account for dependency in effect sizes (Assink & Wibbelink, 2016).

Three-level Meta-Analyses

When primary studies of a meta-analysis report multiple effect sizes, it is important to consider that effect sizes from the same study may correlate and that dependency among effect sizes may exist (Fernández-Castilla et al., 2020). A common case of dependence is the occurrence of multiple observed measures within the same study (Van den Noortgate et al., 2015). A clinical trial might, for example, test the effect of psychedelic therapy on two different constructs, such as anxiety and depression, or might use multiple assessment tools for the same construct, e.g., two different depression inventories. According to Assink and Wibbelink (2016), there are several prevalent methods for handling the dependency of effect sizes, ranging from simply ignoring the dependency, averaging them into one effect size, or selecting only one effect size per study. Selecting only one effect size per study would imply homogeneity of effect sizes within studies. This would be a questionable assumption in most cases (Cheung, 2015). Ignoring the dependency can lead to an overlap of information resulting in the underestimation of standard errors. Type I error inflation is a potential consequence of this (Van den Noortgate et al., 2015).

By applying the statistical model of a multilevel meta-analysis, all relevant information can be preserved, and maximal statistical power can be achieved (Assink & Wibbelink, 2016). The most common type of multilevel meta-analysis is the three-level meta-analysis (Assink & Wibbelink, 2016). This model considers three different variance components of the extracted effect sizes over the three levels: (1) sampling variance at level 1, (2) variance between effect sizes within the same primary study at level 2, and (3) variance between multiple primary studies at level 3. Thus, effect sizes can vary between participants (level 1), outcomes (level 2),

and studies (level 3). While conventional meta-analyses only allow for examining differences between studies (between-study heterogeneity), the three-level approach additionally allows for examining differences in outcomes within studies (within-study heterogeneity, Assink & Wibbelink, 2016).

In the following chapter, I will describe my objectives for this present multilevel meta-analysis on the effects of psilocybin- and MDMA-assisted psychotherapy.

Objectives

Even though the meta-analysis by Leudesdorff (2021) already examined the pre-post summary effects of psychedelic-assisted psychotherapy, this present meta-analysis conceptionally differs by three major aspects: (1) while the meta-analysis by Leudesdorff (2021) examined the effects of multiple psychedelic substances (ayahuasca, ibogaine, psilocybin, LSD, and MDMA), this present meta-analysis focusses on the most promising FDA-designated 'breakthrough therapies' psilocybin and MDMA. To my knowledge, no previous meta-analysis has done that before. (2) Leudesdorff (2021) used the conventional meta-analytical model to analyze the pre-post summary effects, while this present meta-analysis uses the novel three-level meta-analytic approach to account for dependency in effect sizes. To my knowledge, this is also the first time in the field of psychedelic research. (3) Even though only two years lie in between the unpublished master thesis of Leudesdorff (2021) and this present meta-analysis, the psychedelic research field is fast-growing, and very recent publications (Mitchell et al., 2021) that are not included in the meta-analysis of Leudesdorff (2021) have impacted the field immensely. For these reasons, this meta-analysis can be perceived as a potential gain for the field of psychedelic research.

This present master thesis will analyze the potential therapeutic effects of psilocybin- and MDMA-assisted psychotherapy using a three-level meta-analysis. My research objectives are to test the overall short-term as well as the overall long-term outcomes. Because psilocybin- and MDMA-assisted psychotherapies work in different ways and are applied to different patient groups (usually psilocybin for depression and MDMA for PTSD), it will be tested, whether the type of substance and the type of disorder are potential moderators of the short-term summary effect. Furthermore, existing psychedelic research studies vary in their study design and quality. Another objective is therefore to examine whether the summary effect is moderated by study design or study quality.

The following research questions and hypothesis for this meta-analysis are:

1. Objective: Short-term Summary Effect (Acute Effects)

Do symptoms of psychological disorders change shortly after psychedelic therapy? What is the pre-post summary effect of psilocybin and MDMA on mental disorders?

H0: The short-term summary effect is not significant, indicating that the symptoms of mental disorders do not change after psychedelic therapy.

H0: Hedges' $g = 0.0$; $p \geq .05$

H1: The short-term summary effect is significant, indicating that the symptoms of mental disorders do change after psychedelic therapy.

H1: Hedges' $g \neq 0.0$; $p < .05$

2. Objective: Long-term Summary Effect (Persisting Effects)

Are there persisting changes in symptoms of psychological disorders after psychedelic therapy? What is the pre-follow-up summary effect of psilocybin and MDMA on mental disorders?

H0: The long-term summary effect is not significant, indicating that there are no persisting changes in the symptoms of mental disorders after psychedelic therapy.

H0: Hedges' $g = 0.0$; $p \geq .05$

H1: The long-term summary effect is significant, indicating that there are persisting changes in the symptoms of mental disorders after psychedelic therapy.

H1: Hedges' $g \neq 0.0$; $p < .05$

3. Objective: Moderation of Acute Effects by 'Substance'

Is the pre-post summary effect of psychedelic therapy for psychological disorders moderated by the type of substance (psilocybin and MDMA)?

H0: The test of the moderator 'substance' is not significant, indicating that the changes in symptoms of mental disorders after psychedelic therapy are not influenced by the type of substance (psilocybin and MDMA).

H0: $p \geq .05$

H1: The test of the moderator 'substance' is significant, indicating that the changes in symptoms of mental disorders after psychedelic therapy are influenced by the type of substance (psilocybin and MDMA).

H1: $p < .05$

4. Objective: Moderation of Acute Effects by 'Disorder'

Is the pre-post summary effect of psychedelic therapy for psychological disorders moderated by the type of disorder?

H0: The test of the moderator 'disorder' is not significant, indicating that the changes in symptoms of mental disorders after psychedelic therapy are not influenced by the type of disorder.

H0: $p \geq .05$

H1: The test of the moderator 'disorder' is significant, indicating that the changes in symptoms of mental disorders after psychedelic therapy are influenced by the type of disorder.

H1: $p < .05$

5. Objective: Moderation of Acute Effects by 'Study Design'

Is the pre-post summary effect of psychedelic therapy for psychological disorders moderated by the study design?

H0: The test of the moderator 'study design' is not significant, indicating that the changes in symptoms of mental disorders after psychedelic therapy are not influenced by the study design.

H0: $p \geq .05$

H1: The test of the moderator 'study design' is significant, indicating that the changes in symptoms of mental disorders after psychedelic therapy are influenced by the study design.

H1: $p < .05$

6. Objective: Moderation of Acute Effects by 'Study Quality'

Is the pre-post summary effect of psychedelic therapy for psychological disorders moderated by the quality of the study?

H0: The test of the moderator 'study quality' is not significant, indicating that the changes in symptoms of mental disorders after psychedelic therapy are not influenced by study quality.

H0: $p \geq .05$

H1: The test of the moderator 'study quality' is significant, indicating that the changes in symptoms of mental disorders after psychedelic therapy are influenced by study quality.

H1: $p < .05$

Methods

Protocol and Registration

This study was planned and conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses **(**PRISMA) guidelines (Moher et al., 2009) and preregistered through the international prospective register of systematic reviews (PROSPERO, https://www.crd.york.ac.uk/prospero, ID: CRD42022303022, 17/02/2022).

Eligibility Criteria

Eligibility criteria were defined according to the 'PICO model' as specified in the Cochrane Handbook for systematic reviews (McKenzie et al., 2021). This acronym translates to population, intervention, comparators, and outcomes and provides a set of categories according to which eligibility criteria can be specified. (McKenzie et al., 2021).

For study populations, I searched for clinical trials with patient samples diagnosed with anxiety and depression caused by life-threatening illness (LTI, e.g., cancer patients), unipolar depression, like major depressive disorder (MDD), PTSD, anxiety disorder (AD) and substance use disorder (SUD), like alcohol use disorder (AUD). These disorders were considered eligible for the analysis as prior research has demonstrated their susceptibility to the effects of psychedelic therapy (Bahji et al., 2020; Dos Santos et al., 2016; Goldberg et al., 2020; Illingworth et al., 2021; Luoma et al., 2020; Tedesco et al., 2021).

With respect to the intervention, all clinical trials involving the use of psychoactive substances (psilocybin, MDMA, LSD, DMT, and ayahuasca) were initially considered eligible. I started the literature search with a broader scope of psychedelic trials to avoid the current problem of redundant and incomplete meta-analyses (Ioannidis, 2016). However, for some of the interventions, only a small number of studies were found. To achieve a sufficient statistical power of the meta-analysis, interventions were consequently narrowed down to the two substances most reported on - psilocybin and MDMA.

As opposed to previous meta-analyses in the field, no limitations were set concerning study designs, to ensure a sufficiently high sample size and statistical power. Randomized controlled trials (RCTs), open-label, and cross-over studies were considered eligible. Therefore, both kinds of comparisons, between-subjects, and within-subjects, were included.

Finally, studies were required to measure at least one psychological outcome (score or professional rating on a symptom scale) over at least two time points, including before and after the treatment with psychedelic substances, to analyze the change in symptoms. In case of

missing information, corresponding authors were contacted. No restriction was placed on publication date, gender, ethnic background, age, language, or publication status (published or unpublished).

Information Sources

The basis of a robust meta-analysis is a thorough literature search (Card, 2012). Following Card's (2012) guidelines, I used different kinds of information sources. In addition to searching databases, backward searches were conducted and research registries, as well as websites of research institutions, were searched to examine suitable studies. Online databases were chosen for diversity in their functional area and to allow for both, published and unpublished articles.

Five databases were searched for published literature, including PsycINFO, PubMed, Scopus, Web of Science, and Google Scholar. Furthermore, three databases were searched for grey literature, including ProQuest Dissertations & Theses Global and the two research registries ClinicalTrials.gov and WHO's International Clinical Trials Registry Platform (ICTRP). Whenever given, I set a filter for clinical trials, but no restriction in time or language. The database search was conducted between October 13th and 25th, 2021 and Search alerts were set. The last inclusion based on search alerts was made on the 31st of August 2022.

Additionally, I conducted a database search of existing meta-analyses and systematic reviews. The samples of the meta-analyses were subsequently hand-searched for additional studies (backward search). Finally, I manually searched the research sections of the psychedelic research institutes MAPS and Usona for eligible clinical trials and reviewed the most recent psychedelic studies presented at the annual Horizons Conference in New York, in December 2021 for this sample.

Search

Individual search strings for each database were created. All search strings included the same keywords that were systematically deduced from the predefined eligibility criteria.

The keywords for the interventions ("Psilocybin", "MDMA", …) and their variants were first combined with the Boolean operator OR and then combined with the Boolean operator AND with the connected keywords for the population ("Anxiety", "Depression", "PTSD", …). When given, filters were applied to exclude certain document types, like systematic reviews. All search strings for each database with search dates and search results (hits) for both, primary studies (PS) and meta-analyses (MA) are displayed in Table 1.

Table 1. *Database Searches for Primary Studies and Meta-Analysis*

Databases for Published Articles

Database	Search String	Hits (PS)	Hits (MA)	Comments
PsychINFO	(Psilocybin or MDMA or 3,4-methylenedioxymethamphetamine or ayahuasca or DMT or lysergic acid diethylamide or LSD or psychedelic*).mp. AND (Anxiety or depression or post*traumatic stress disorder or PTSD).mp. AND (Treatment or therap*).mp. Filter: "therapy (best balance of sensitivity and specificity)"	125	12	Last search 13.10.2021 *(Search alert set)*
Pubmed	(Psilocybin OR MDMA OR 3,4-methylenedioxymethamphetamine OR ayahuasca OR DMT OR lysergic acid diethylamide OR LSD OR psychedelic*) [Title/Abstract] AND (Anxiety OR depression OR dependence OR substance use OR OCD OR posttraumatic stress disorder OR PTSD) [Title/Abstract] AND (Treatment OR therap*) [Title/Abstract] Filter: Clinical Trial	77	13	Last search 13.10.2021
Scopus	(psilocybin) OR (mdma) OR (3,4-methylenedioxymethamphetamine) OR (ayahuasca) OR (dmt) OR (lysergic acid diethylamide) OR (lsd) OR (psychedelic*) [TITLE-ABS-KEY] AND (anxiety) OR (depression) OR (dependence) OR (substance use) OR (ocd) OR (post*traumatic stress disorder) OR (ptsd) [TITLE-ABS-KEY] AND (treatment) OR (therap*) [TITLE-ABS-KEY] AND (clinical trial) [TITLE-ABS-KEY] Filter: Document Type → Article	373	46	Last search 19.10.2021 *(Search alert set)*
Web of Science	(Psilocybin OR MDMA OR 3,4-methylenedioxymethamphetamine OR ayahuasca OR DMT OR lysergic acid diethylamide OR LSD OR psychedelic*) [All Fields] AND (Anxiety OR depression OR dependence OR substance use OR OCD OR posttraumatic stress disorder OR PTSD) [All Fields] AND (treatment) OR (therap*) [All Fields] AND (clinical trial) [All Fields] Filter: Exclude Document Type → Reviews	158	23	Last search 19.10.2021 *(Search alert set)*
Google Scholar	*Advanced Search Google:* *Articles with all of the words:* Psilocybin *Articles with at least one of the words:* Clinical trial (In the title of the article) *Articles with all of the words:* MDMA *Articles with at least one of the words:* Clinical trial (In the title of the article) *Articles with all of the words:* LSD *Articles with at least one of the words:* Clinical trial (In the title of the article) *Articles with all of the words:* Ayahuasca *Articles with at least one of the words:* Treatment Therapy (In the title of the article)	264	-	Last search 25.10.2021

Databases for Unpublished Articles (Dissertations & Research Registries)

Database	Search String	Hits (PS)	Hits (MA)	Comments
ProQuest Dissertations & Theses Global	(Psilocybin OR MDMA OR 3,4-methylenedioxymethamphetamine OR ayahuasca OR DMT OR lysergic acid diethylamide OR LSD OR psychedelic*) [Abstract] AND (Anxiety OR depression OR dependence OR substance use OR OCD OR posttraumatic stress disorder OR PTSD) [Abstract] AND (treatment) OR (therap*) [Abstract]	99	-	Last search 19.10.2021 (*Search alert set)*
ClinicalTrial.gov	Condition: Anxiety OR depression OR dependence OR OCD OR PTSD AND	16	-	Last search 25.10.2021

	Intervention: Psilocybin OR MDMA OR ayahuasca OR DMT OR LSD OR psychedelic* AND Filter: Interventional (Clinical Trial) Filter: Has Results			*(Search alert set)*
WHO's Clinical Trials Registry Platform	(Anxiety OR depression OR dependence OR OCD OR PTSD) AND (Psilocybin OR MDMA OR ayahuasca OR DMT OR LSD OR psychedelic*) Filter: With results only	10	-	Last search 25.10.2021 *(Search alert set)*

Note. 1122 Hits of Primary Studies (PS) and 94 hits of Meta-analyses (MA) were found and exported to Endnote X9 (Clarivate, 2021).

Study Selection

All primary studies were saved and screened in the reference management software EndNote X9 (Clarivate, 2021). First, citation duplicates were automatically detected and removed. After that, I manually screened titles and abstracts for inclusion. Exclusions were made based on study type, population, intervention, and document type. After the identification of relevant studies, missing full texts were asked for retrieval. Finally, I screened all full texts retrieved according to the prespecified inclusion and exclusion criteria for my sample.

Data Collection Process

As a first step, full texts of the articles were scanned to obtain relevant information. When data for calculating the effect sizes (means and standard deviations) were not provided in the tables of the studies, supplementary materials or other online resources of the research teams were examined. In some cases, it was necessary to contact the corresponding author because only limited information was reported, or data was not shared at all. In three cases (Grob et al., 2011; Mitchell et al., 2021; Ross et al., 2016) the authors did not respond. Therefore, I used the web-based tool WebPlotDigitizer by Rohatgi (2020) to extract numerical data from plots, graphs, and images of the reports (see Table 1S, supplementary material).

All data items were first coded by myself and then independently coded again by my colleague Stefanie Ochmanek. After this, an interrater reliability (IRR) was calculated using the Agreeableness rate (AR). The AR is the most widely used index in research synthesis, which is computationally simple and intuitively interpretable (Cooper et al., 2019). The AR is the result of the number of observations agreed upon (763) divided by the total number of observations (777). The IRR for this data collection process was very high (AR = 98,2%), which means that the two coders agreed on 98,2% of all variables coded (Cooper et al., 2019). The IRR for the Risk of Bias Assessments was calculated separately and is reported in the chapter 'Risk of Bias in Individual Studies'.

Data Items: Coding

In accordance with the standards provided by Cooper et al. (2019), data items (variables) were generated to address the research questions and to operate all preregistered statistical analyses. I created a coding guide and three different coding sheets containing study characteristics, necessary information to calculate effect sizes, and data items to conduct moderator analyses.

All essential study characteristics, such as the population, intervention, study design, dosing, and placebo, were coded. Descriptive data, like sample sizes, mean age of participants, and percentage of females in the sample, were also collected. I additionally coded the occurrence of adverse events, reasons for dropouts, and whether a peak experience was measured or not. Dropouts were defined as participants leaving the study after and not before the first dosing session because this time of exit can be attributed to a potential negative reaction to the drug administration.

To calculate the change in symptoms of psychological disorders (effects), I chose the assessment tools based on both, the best qualitative fit to the constructs, and the most overlap of all studies which resulted in two or three (depending on applicability) instruments per study. Means (*M*), standard deviations (*SD*), and sample sizes (*N*) were collected at three assessment time points: Baseline, first time point after all dosing sessions for the short-term effect, and – when applicable – follow-up timepoint after all dosing sessions for the long-term effect. I chose time points based on most available data and most overlap of all studies. In the case of two-arm trials, only data from the treatment groups, not the control groups were included. This was done to focus on within-subjects (pre-post) effects.

Moderators were defined considering sensible factors that can be influential to the overall therapeutic effect. The following preregistered variables were coded as moderators: (1) ‘substance’ – due to the inclusion of different psychedelic substances in the sample, (2) ‘disorder’ – because studies varied across clinical populations, (3) ‘study design’ – because between-subjects and within-subjects study designs were included and (4) ‘study quality’ – based on the risk of bias assessments. Additional factors were coded as potential moderators: (5) ‘instrument’ – determining whether the instrument was self-report or clinician-administered, (6) ‘dosing’ – of the substance, and (7) ‘peak experience effect’ – indicating whether there was a significant peak experience measured or not. Dosing of the substance was rated according to classifications from primary studies (see Table 1S, supplementary material). The study was coded as ‘high’ when high dosing was used at least in one condition of the study.

The study was coded as ‘moderate’ when no high dosing was used in any condition. Low dosing was only used as a control and not in experimental groups.

Risk of Bias in Individual Studies: Risk of Bias Tool

To assess study quality, Cochrane’s risk-of-bias tool for randomized trials ([RoB] Higgins et al., 2011) was used. Five domains of bias were considered: (1) selection bias (random sequence generation, allocation concealment), (2) performance bias (blinding of participants and personnel), (3) detection bias (blinding of outcome assessors), (4) attrition bias (incomplete outcome data), and (5) reporting bias (selective reporting). For each domain, an evaluation of low, high, or unclear risk of bias was made, and a record of the eventual assessment was taken. Each study was rated independently by me and my colleague Stefanie Ochmanek with a high AR of 89.3%. The risk of bias was rated conservatively to control for often highly subjective psychedelic experiences.

The preregistered risk of bias measure Newcastle-Ottawa Scale (Ottawa Hospital Research Institute, 2011) for open-label studies was ultimately excluded from the quality assessment because my open-label studies did not fit its categories (case-control or cohort studies). As a solution, I used Cochrane’s RoB for all my studies, RCT and open-label, as a previous meta-analysis did (Goldberg et al., 2020). It is important to note that the interpretation of the tool is therefore limited because it is not designed for the rating of single-arm studies.

For the visualization of the study quality ratings, I used the online tool ‘Robvis’ to create a traffic light plot (*Risk of Bias Tools*, n.d.) and modified the plot to fit the five domains of Cochrane’s RoB.

Summary Measures: Hedges’ g

All analyses were performed using R Statistical Software (v3.6.3; R Development Core Team, 2020). To analyze the change in symptoms of psychological disorders (effects), the mean differences between assessment outcomes before and after psychedelic treatment were calculated. I chose a within-subjects comparison because most studies provided data for pre-post effects only and, therefore, more studies could be included, which eventually allowed for higher statistical power.

As an effect size measure, I chose Hedges’ *g*. It can be determined using Cohen’s *d* multiplied by a correction factor *J* for small samples and represents the standardized mean difference between two groups or time points (Borenstein et al., 2021). Hedges’ *g* was chosen because *d* has a slight bias to overestimate the effect in small samples, and all included primary

studies used relatively small sample sizes. I applied the formula provided by Borenstein et al. (2021, p. 26 & 27) to calculate the effect sizes for each relevant assessment tool per primary study in this sample. The single short-term and long-term effects for selected instruments were consequently calculated using the means of within-group pre–post and pre–follow–up scores and their standard deviation of differences as the measure of dispersion.

Synthesis of Results: Three-Level Meta-Analytic Model

Short-term and Long-term Summary Effects

The summary effect is the main outcome of a meta-analysis. All Hedges' *g* values are summarized to represent the overall effect of all included primary studies. To answer the first and second research questions, the short-term and long-term summary effects were calculated using a three-level meta-analytic approach, following the tutorial by Assink and Wibbelink (2016).

A within-group pre–post and a pre–follow–up overall Hedges' *g* was computed using the rma.mv function of the metafor package in R (Viechtbauer, 2015). Random effects were assumed and the rma.mv function was invoked with the random argument. The Knapp and Hartung (2003) adjustment was applied to the analyses to reduce the problem of increased significant results when testing parameters using the z-distribution, which is a default for the rma.function. REstricted Maximum Likelihood estimation method (REML) was applied for estimating all parameters.

Heterogeneity of Short-term and Long-term Summary Effects

One advantage of the three-level meta-analytic model is, that it considers three different variance components: (1) The sampling variance based on participants on level one, (2) the variance between effect sizes of outcomes within the studies on level two, and (3) the variance of effect sizes between the studies on level three. Therefore, this model allows controlling for two different kinds of heterogeneity: the within-study heterogeneity on level two and the between-study heterogeneity on level three (Assink & Wibbelink, 2016).

To test the significance of these two heterogeneities, two one-sided log-likelihood ratio tests were conducted for each summary effect short and long. As the likelihood tests are by default two-sided in R, the p-values were divided by two. Within these tests, the fit of the three-level model was compared to the fit of a two-level model. Finally, I tested the distribution of variance over the three levels of the meta-analytic model (I^2). I^2 is a descriptive proportion that reflects variance in effects rather than sampling error (Borenstein et al., 2021).

Forest Plots of Short-term and Long-term Summary Effects

For the visualization of the summary effects short and long, I created two forest plots in R using the metaviz 0.3.1 package (Comprehensive R Archive Network [CRAN], 2020). The forest plot portrays the direct relationship between the precision of a study and the extent to which that study contributes to the final estimation of the overall effect (Fernández-Castilla et al., 2020). The plots were modeled with random effects models instead of multilevel models to provide a more intuitive representation.

Moderator-analyses

Moderator-analyses assess the relationship between relevant covariates on the levels of studies and the effect sizes (Borenstein et al., 2021). To determine whether a moderating effect of potential moderators included in the model is significant, omnibus *F*-tests were performed using the the rma.mv function of the metafor package in R (Viechtbauer, 2015). The moderator analyses were conducted to determine whether the predefined moderators (1) 'substance', (2) 'disorder', (3) 'study design' and (4) 'study quality' moderate the short-term summary effect significantly to answer the third, fourth, fifth and sixth research questions, respectively. Additionally, the potential moderators (5) 'instrument', (6) 'dosing', and (7) 'peak experience effect' were tested for moderation.

Testing Significance and Contrasts of Moderators 'Substance' and 'Disorder'

The moderator 'substance' was tested as a dichotomous moderator of the short-term summary effect (0 = psilocybin; 1 = MDMA) with contrasts to answer the third research question. A contrast test of a moderator can determine a separate mean effect size for each category of the moderator (Assink & Wibbelink, 2016). Taking for example the moderator 'substance', a contrast test can determine, whether the category 'psilocybin' is significantly moderating the summary effect in comparison to the category 'MDMA' and what the separate summary effects of psilocybin and MDMA studies are. I was especially interested to calculate separate summary effects for the categories of the moderator 'substance' to see, which of the two substances has a higher treatment effect. Therefore, the moderator 'substance' was tested for contrasts to get separate summary effects for psilocybin and MDMA. The results of the separate summary effects can be viewed in the Figures 3 and 4.

The moderator 'disorder' was also tested with contrasts as a categorical moderator of the short-term summary effect with five categories (0 = resistant MDD; 1 = depression and anxiety in patients with LTI; 2 = PTSD; 3 = AUD; 4 = social anxiety) to answer the fourth research question.

Significance of Moderators 'Study Design', and 'Study Quality'

Additional moderator analyses were performed with the dichotomous moderator 'study design' (0 = between subjects; 1 = within subjects) and 'study quality' (0 = low study quality; 1 = high study quality) to answer the fifth and sixth research questions. I inversed the coding of the risk of bias assessment for 'study quality' for a more intuitive interpretation (0: low study quality = high RoB, and 1: high study quality = low RoB). The studies with unclear RoB were ultimately categorized as low study quality to simplify the analyses. This rather conservative approach was taken because the withholding of information usually can be interpreted that measures were not taken and therefore a high risk of bias.

Exploratory: Significance of Moderators 'Instrument', 'Dosing', and 'Peak Experience Effect'

The dichotomous moderators 'instrument' (0 = self-report; 1 = clinician-administered), 'dosing' (0 = moderate dosing; 1 = high dosing), and 'peak experience effect' (0 = no effect of peak experience on the outcome; 1 = effect of peak experience on the outcome) were tested exploratively. 'Peak experience effect' was only measured in the seven psilocybin studies.

Risk of Bias across Studies: Publication Bias Assessment

According to Borenstein et al. (2021), studies that report large effect sizes are more likely to be published than studies with small effect sizes. If the sample of primary studies is systematically biased, the computed summary effect by the meta-analysis will reflect this bias. Publication bias is especially common in small-sampled studies (Borenstein et al., 2021).

The first measure that was taken to prevent potential bias in this meta-analysis, was to conduct a thorough literature search with different information sources of not published reports (see chapter information sources). For the visualization of the relationship between effect size and sample size in my sample, a funnel plot was conducted in R using the metaviz 0.3.1 package (CRAN, 2020). Because the interpretation of a funnel plot is largely subjective (Borenstein et al., 2021), Egger's regression test (Egger et al., 1997) was performed additionally to statistically test for publication bias. This test measures if there is a significant relationship between effect size and sample size.

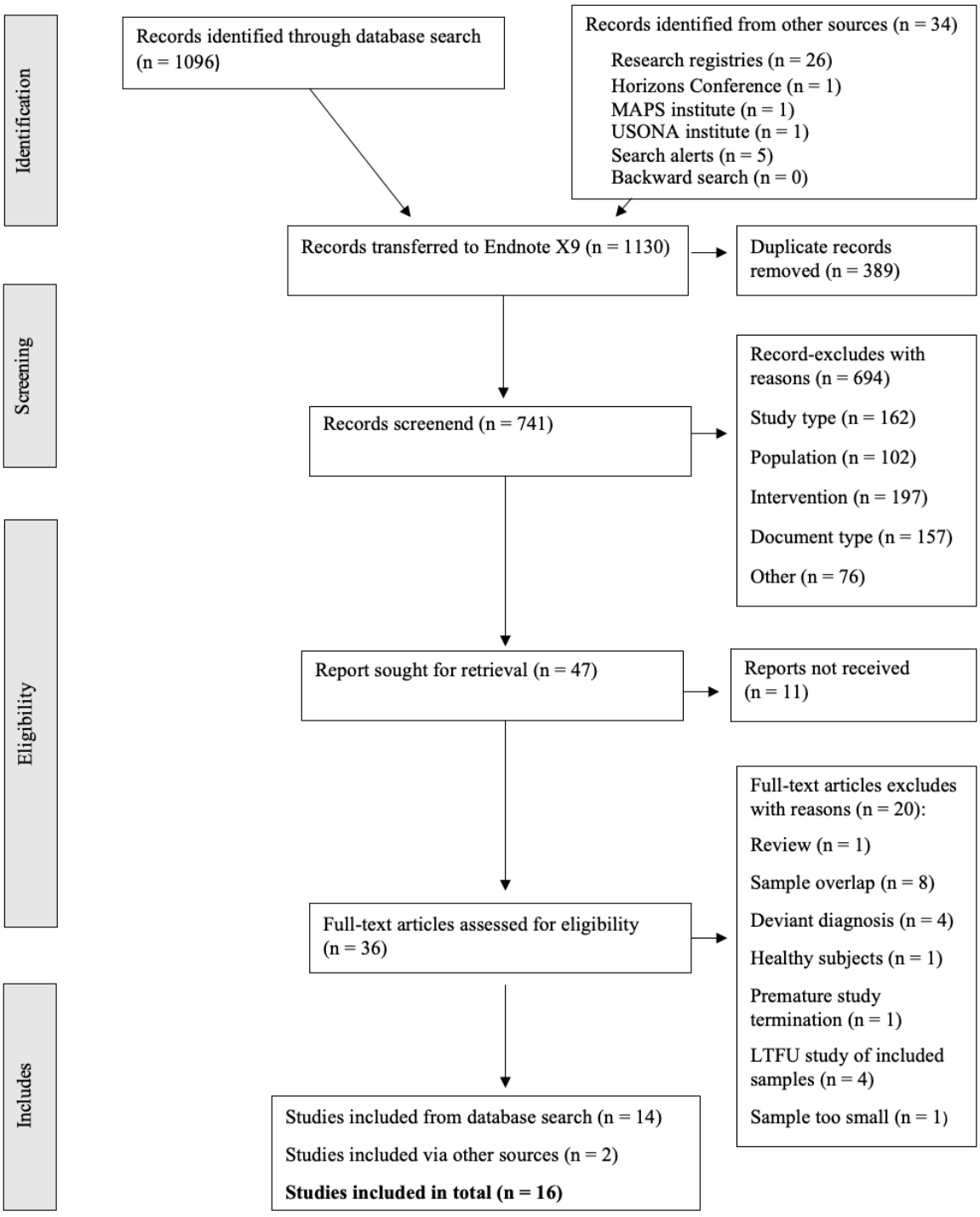
Results

In the following chapter, all results of the literature search and statistical analyses will be reported according to the standards of the PRISMA statement (Moher et al., 2009).

Study Selection: PRISMA Flow Diagram

The process of study selection is displayed in Figure 1. All 1122 References from the database searches and 8 Studies from searches using other methods were exported to the reference management software EndNote X9 (Clarivate, 2021) and thoroughly screened. All exclusion criteria in the selection process of the primary studies can be seen in Figure 1. Special attention was paid to not include studies with overlapping samples with other included studies. Ultimately, 16 Primary Studies with unique samples (N = 463) were included in the meta-analyses.

Figure 1. *Adapted PRISMA 2020 Flow Diagram for Clinical Studies of Psychedelic Therapy*



Note. Modified PRISMA flow diagram by Page et al. (2021) including searches of databases, research registries, research institutes, research conferences, search alerts, and backward searches. LTFU = long-term follow-up.

Study Characteristics

Characteristics of the 16 Primary Studies are reported in Table 2. Seven clinical trials administering psilocybin (N = 208) and nine clinical trials administering MDMA (N = 255) were included in this meta-analysis. Clinical diagnoses of patient groups treated varied across primary studies. Studies enrolled patients suffering from resistant PTSD (7 studies), anxiety and depression in patients with LTI (e.g. cancer, 4 studies), resistant MDD (3 studies), AUD (1 study), and social anxiety in autistic adults (1 study). While ten studies were already included in the unpublished meta-analysis by Leudesdorff (2021), six 'new' studies, published in 2000 - 2021, were included in this present meta-analysis.

Primary Endpoints

As primary endpoints, mean scores of listed instruments (Table 2) assessing the severity of symptoms of psychological disorders were extracted before and after treatment with psychedelic substances. Three instruments assess symptoms of PTSD: the Clinician-Administered PTSD Scale (CAPS & CAPS IV, Blake et al., 1995; CAPS 5, Weathers et al., 2013), the Impact of Event Scale-Revised (IES-R, Weiss & Marmar, 1996), and the Posttraumatic Diagnostic Scale (PDS, Foa et al., 1997). Two instruments assess depressive symptoms: the Beck Depression Inventory (BDI, Beck et al., 1961; BDI-II, Beck et al., 1996) and the Quick Inventory of Depressive Symptomatology (QIDS SR-16, Rush et al., 2003). Two instruments assess symptoms of AUD: the Penn Alcohol Craving Scale (PACS, Flannery et al., 1999) and the Temptation Subscale of the Alcohol Abstinence Self-Efficacy Scale (AASE, DiClemente et al., 1994). One instrument assesses symptoms of anxiety: the State-Trait Anxiety Inventory (STAI, Spielberger et al., 1983). In all instruments, decreases in values indicate symptom reduction. All clinical assessment tools were self-report questionnaires, but the CAPS, CAPS-IV and CAPS-5, which are structured interviews designed to be administered by clinicians and clinical researchers. To account for this difference in measurement, the categories 'self-report' and 'clinician-administered' were considered in the explorative moderator 'instrument' (see chapter 'moderators').

More detailed information on study characteristics, i.e. the exact dosing, the course of the clinical trial, how data was won, chosen assessment tools and assessment time points, effects of peak experiences, dropout reasons, and adverse effects can be seen in Table 1S in the supplementary material. It is important to note that the latest published MDMA study, which is included in this sample, is the first-ever 'phase 3' clinical trial on psychedelic substances with a sample of 90 participants (Mitchell et al., 2021). This study is regarded as a milestone in

psychedelic research (MAPS, 2022), and its inclusion increases the statistical power of the present meta-analysis.

Table 2. *List of 16 Primary Studies Included in Three-Level Meta-Analysis*

7 Psilocybin Studies (*N* = 208)						
Author	**Year**	**Intervention**	**Population**	***N***	**Study Design**	**Endpoints**
Grob et al.*	2011	Psilocybin	Anxiety & depression in cancer patients	12	RCT with Placebo (Niacin)	BDI STAI-S STAI-T
Bogenschutz et al.*	2015	Psilocybin	AUD	10	Open Label	PACS AASE
Griffiths et al.*	2016	Psilocybin	Anxiety & depression in cancer patients	51	RCT (cross-over) low dose as Placebo	BDI STAI-S STAI-T
Ross et al.*	2016	Psilocybin	Anxiety & depression in cancer patients	29	RCT (cross-over) with Placebo (Niacin)	BDI STAI-S STAI-T
Carhart-Harris et al.*	2017	Psilocybin	Resistant MDD	20	Open Label	BDI STAI-T
Davis et al.	2021	Psilocybin	Resistant MDD	27	RCT delayed treatment as control	BDI-II STAI-S STAI-T
Carhart-Harris et al.	2021	Psilocybin	Resistant MDD	59	RCT low dose as Placebo	QIDS-SR-16 BDI – 1A
9 MDMA Studies (*N* = 255)						
Author	**Year**	**Intervention**	**Population**	***N***	**Study Design**	**Endpoints**
Mithoefer et al.*	2010	MDMA	Resistant PTSD	20	RCT with Placebo (Lactose)	CAPS IES-R
Oehen et al.*	2013	MDMA	Resistant PTSD	12	RCT low dose as Placebo	CAPS PDS
Danforth et al.*	2018	MDMA	Social anxiety in autistic adults	12	RCT with Placebo (Lactose)	BDI-II STAI-S STAI-T

Mithoefer et al.*	2018	MDMA	Resistant PTSD	26	RCT low dose as Placebo	CAPS-IV BDI-II
Ot'alora et al.*	2018	MDMA	Resistant PTSD	28	RCT low dose as Placebo	CAPS-IV BDI-II
Monson et al.	2020	MDMA & CBCT	Resistant PTSD	12	Open Label	CAPS-5 BDI-II
Wolfson et al.	2020	MDMA	Anxiety in patients with LTI	18	RCT With Placebo (Lactose)	STAI-T STAI-S BDI-II
Wang et al.	2021	MDMA	Resistant PTSD	37	Open Label	CAPS-5
Mitchell et al.	2021	MDMA	Resistant PTSD	90	RCT inert Placebo	CAPS-5 BDI-II

Note. Studies in the sample of Leudesdorff (2021) are marked with *; *N* = sample size; RCT = randomized controlled trial; AUD = alcohol use disorder; MDD = major depressive disorder; MDMA = 3,4-Methylenedioxymethamphetamine; PTSD = post-traumatic stress disorder; CBCT = cognitive-behavioral conjoint therapy; LTI = life-threatening illness; BDI = Beck Depression Inventory; STAI-S = State-Trait Anxiety Inventory-State; STAI-T = State-Trait Anxiety Inventory -Trait; PACS = Penn Alcohol Craving Score; AASE = Alcohol Abstinence Self-Efficacy Scale; QIDS SR-16 = Quick Inventory of Depressive Symptomatology; CAPS = Clinician-Administered PTSD Scale; IES-R = Impact of Event Scale; PDS = Posttraumatic Diagnostic Scale.

Risk of Bias within Studies

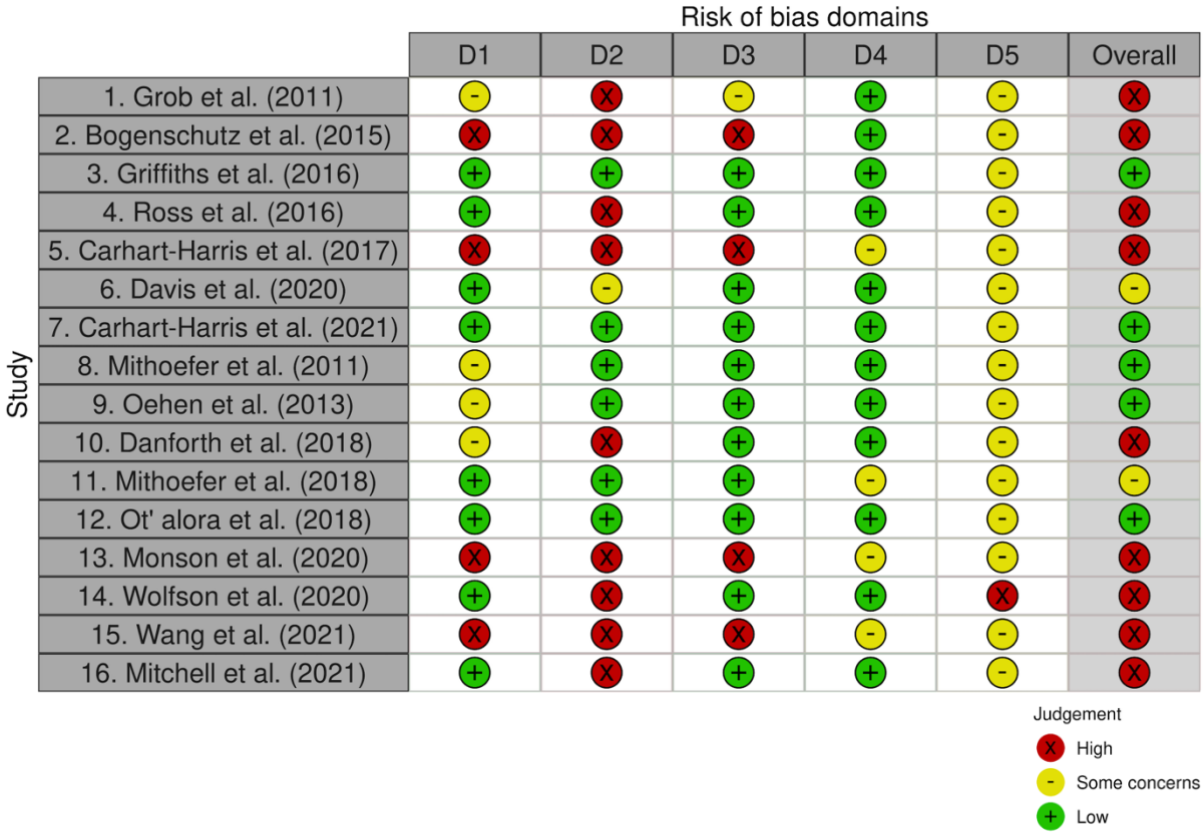
Results of risk of bias assessments with Cochrane's RoB tool (Higgins et al., 2011) are shown in the modified traffic light plot (Figure 2). The risk of bias varied across studies. 'Some concerns' (yellow light) mostly arose when no information about the assessed category could be found. The absence of information was therefore judged as a potential risk of bias.

The risk of bias also varied across domains. Domain 1 'selection bias' was rated as high risk in the four open-label studies (Bogenschutz et al., 2015; Carhart-Harris et al., 2017; Monson et al., 2020; Wang et al., 2021). It is important to note, that this judgment was made because the chosen quality assessment tool is designed for randomized trials (see explanation in chapter Risk of Bias in Individual Studies: Risk of Bias Tool). Domain 2 'performance bias' was rated as high risk, not only when participants and researchers were not blinded, but also when blinding was tested to be unsuccessful. Domain 5 'reporting bias' was predominately

rated with some concerns of bias. It was hard to judge, whether the reporting of outcomes was selective or not because in most cases, outcomes were not prespecified.

For the overall judgment of the risk of bias, three domains were picked to be especially relevant. Following the recommendation of Higgins et al. (2011), 'performance bias', 'detection bias', and 'attrition bias' were valued as relevant domains for highly subjective studies and weighted conservatively. When one of these three domains was judged as 'high risk' or with 'some concerns' of bias, the whole study was ranked as such. Only when all three domains were ranked as 'low risk', the whole study was judged as low risk of bias. This overall judgment was later included as the moderator 'study quality' in subgroup analyses with reversed polarity (low risk of bias = high study quality; high risk of bias & some concerns = low study quality).

Figure 2. *Modified Traffic Light Plot: Risk of Bias Assessments of 16 Primary Studies according to Cochrane's RoB tool (Higgins et al., 2011)*



Note. D1: Selection bias (random sequence generation and allocation concealment)
D2: Performance bias (successful blinding of participants and researchers)
D3: Detection bias (blinding of outcome assessment)
D4: Attrition bias (incomplete outcome data)
D5: Reporting bias (selective reporting)

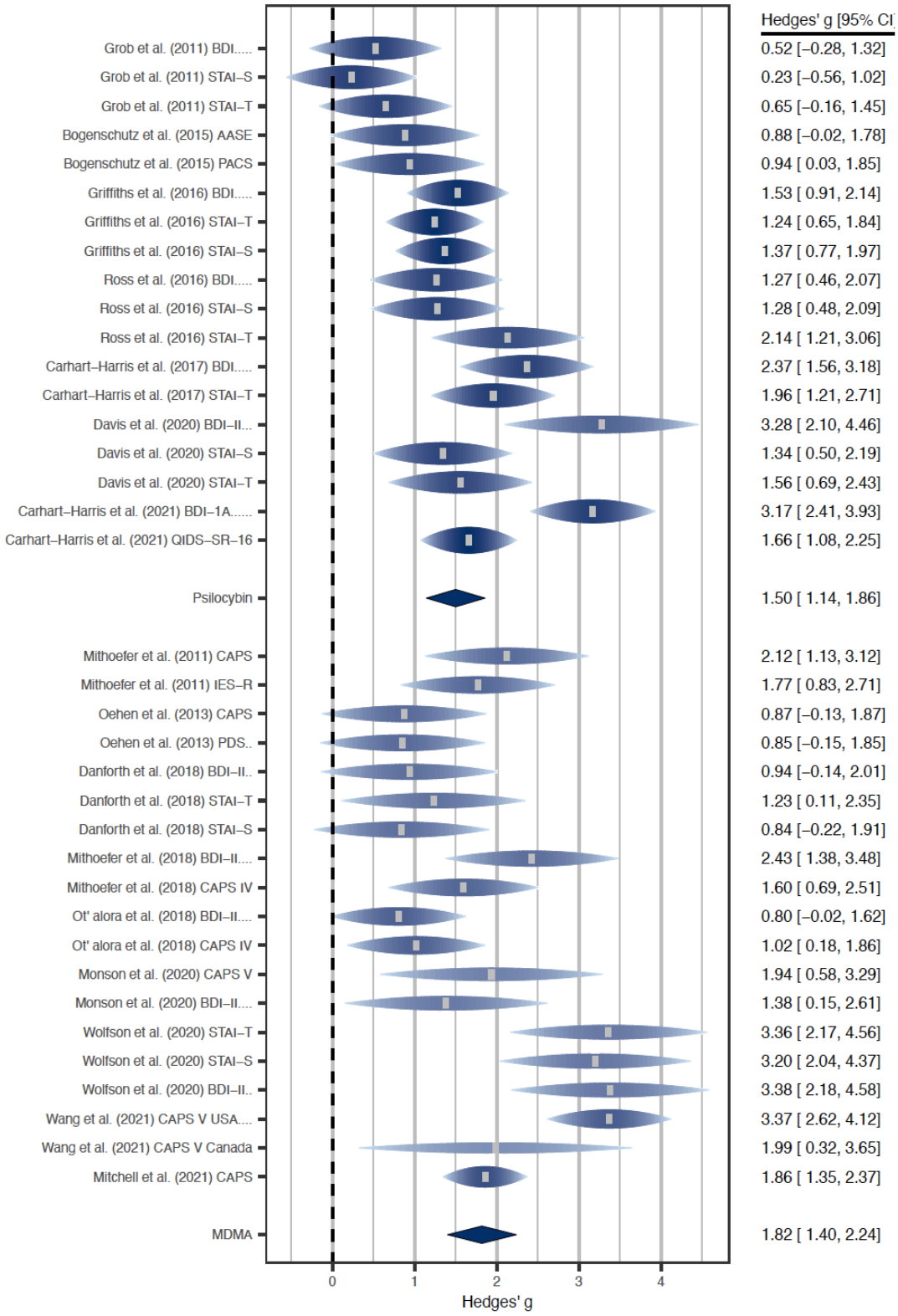
Results of individual Studies

Results of individual studies are presented with rain forest plots in Figures 3 and 4. These forest plots portray the pre–post, and pre–follow–up effects in Hedges'*g* with 95% confidence intervals (CIs) per each assessment for each study. 37 Effect sizes of 16 studies were calculated for the pre–post summary effect and 27 effect sizes of 11 studies were calculated for the pre–follow–up summary effect. Five studies (Davis et al., 2021; Carhart-Harris et al., 2021; Oehen et al., 2013; Ot'alora et al., 2018; Wang et al., 2021) did not conduct follow-up assessments and were therefore not included in the analysis for persisting effects. A positive value of Hedges' *g* represents an increase in symptom reduction and therefore higher efficacy of the drug whereas a negative value represents the opposite. Hedges' *g* values of .2, .5, and .8 can be interpreted as small, moderate, and large effects, respectively (Cohen, 1988).

Forest plots inform about the precision and weight of each instrument used in the study (Fernández-Castilla et al., 2020). Each effect size is depicted with an oval likelihood drop at the horizontal axis. The widths of the 'raindrops' represent the confidence intervals of the effects. At the bottom of the plot, a diamond represents the summary effect across all studies. Compared to a regular forest plot, the plausibility of 'true values' given the observed estimate is visualized by the height and color shading of each likelihood drop (Barrowman & Myers, 2003). Estimates with wide confidence intervals and therefore less weight in the meta-analysis are visually less dominant. It is important to note, that the plots were modeled in R with random effects models instead of multilevel models for a more intuitive representation. The summary effects for each subgroup – psilocybin and MDMA – are therefore minimally less precise with a difference at the second decimal place.

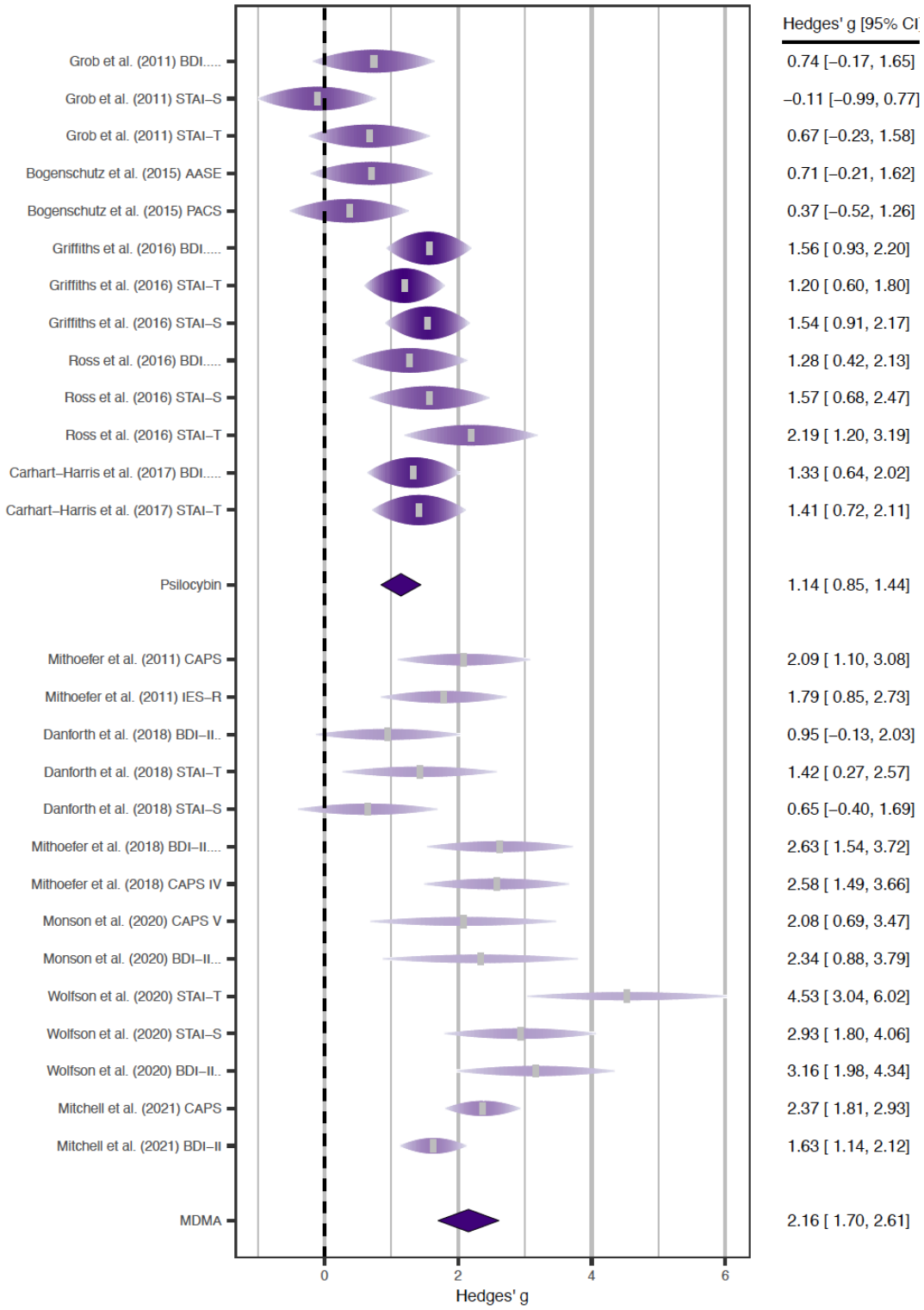
The short-term and long-term summary effects for both substances will be reported in the following chapter to answer the main objectives of this thesis. The overall short-term summary effects for each subgroup – psilocybin and MDMA - will be reported in the chapter 'Moderator-analyses - Contrasts of moderator substance' and discussed in the discussion part of this meta-analysis.

Figure 3. *Rain Forest Plot: Short-term Effects per Assessment across 16 Primary Studies*



Note. Pre–post effects are presented per each assessment for each primary study. Short-term summary effects are shown for each substance – psilocybin and MDMA – separately. BDI = Beck Depression Inventory; STAI = State-Trait Anxiety Inventory; AASE = Alcohol Abstinence Self-Efficacy Scale; PACS = Penn Alcohol Craving Scale; QIDS-SR-16 = Quick Inventory of Depressive Symptomatology; CAPS = Clinician-Administered PTSD Scale; IES-R = Impact of Events Questionnaire; PDS = Posttraumatic Diagnostic Scale.

Figure 4. *Rain Forest Plot: Long-term Effects per Assessment across 11 Primary Studies*



Note. Pre–follow–up effects are presented per instrument for each included study. Long-term Summary effects are shown for each substance – psilocybin and MDMA – separately. BDI = Beck Depression Inventory; STAI = State-Trait Anxiety Inventory; AASE = Alcohol Abstinence Self-Efficacy Scale; PACS = Penn Alcohol Craving Scale; CAPS = Clinician-Administered PTSD Scale; IES-R = Impact of Events Questionnaire.

Synthesis of Results

1. Objective: Short-term Summary Effect (Acute Effects)

To address the first research objective, changes in symptoms of 274 participants (level 1) measured in 37 unique assessments (level 2) in 16 unique studies (level 3) were included in the three-level meta-analyses. Results of the pre-post summary effect of both substances (psilocybin- and MDMA-assisted therapy) across all primary studies are displayed in Table 3. These outcomes indicate a significant and very large reduction in symptoms of psychological disorders after psychedelic therapy (1– 6 weeks) with a Hedges' *g* of 1.684; $p < .001$. This outcome supports the alternative hypothesis (H1: $p < .05$) and the first research question is therefore confirmed.

The test for overall heterogeneity shows significant variation between all effect sizes ($Q(36) = 125.753$, $p < .001$). Significance tests of the within- and between-study heterogeneity in effect sizes are also reported in Table 3. The heterogeneity tests show that the effects are dispersing minimally around the estimated summary effect within one study. This dispersion is not significant. In contrast, the average dispersion of the summary effect between studies is significant and higher. Table 3 also depicts the values of I^2. These results show that 25.96 % of the total variance can be attributed to the within-study sampling variance (level 1), 6.992 % to differences between effect sizes within studies (level 2), and 67.048 % to differences between studies (level 3). These results show that the between-study heterogeneity is substantially higher than the within-study heterogeneity.

Table 3. *Short-term Pre-Post Summary Effect of Psilocybin- & MDMA- Assisted Therapy*

Model Results				
Hedges' *g*	*SE*	*t* (36)	*p*	*CI*
1.684	0.196	8.596	<.001***	1.287 - 2.082
Heterogeneity				
		Sigma	Sqrt	*p*
Within-Study Variance (Level 2)		0.052	0.227	.192
Between-Study Variance (Level 3)		0.495	0.703	.001**
Distribution of Variance (I^2)				
		I^2		
Amount Variance Level 1		25.960		
Amount Variance Level 2		6.992		
Amount Variance Level 3		67.048		

Note. Model Results display outputs of a three-level metanalytic model; *SE* = Standard Error; *CI* = Confidence Interval: heterogeneity was tested with one-sided log-likelihood-ratio tests; I^2 is a descriptive proportion that reflects variance in effects rather than sampling error; Significance codes: 0 '***' 0.001 '**'.

2. Objective: Long-term Summary Effect (Persisting Effects)

To address the second research objective, changes in symptoms of 162 participants (level 1) measured in 27 unique assessments (level 2) in eleven unique studies (level 3) were included in a second three-level meta-analysis to measure the persisting effects. Results of the pre–follow–up summary effect of both substances (psilocybin- and MDMA-assisted therapy) are displayed in Table 4. These outcomes indicate a significant and very large reduction in symptoms of psychological disorders at a follow-up assessment (4 – 12 months) after psychedelic therapy with a Hedges' *g* of 1.650; $p = < .001$. This outcome supports the alternative hypothesis (H1: $p < .05$) and the second research question is therefore confirmed.

The test for overall heterogeneity shows significant variation between all effect sizes ($Q(26) = 85.322$, $p < .001$). Results for the within- and between-study heterogeneity in effect sizes are also shown in Table 4. The effects were dispersed minimally around the estimated summary effect within one study. This dispersion is not significant. Contrary to that, the average dispersion of the summary effect between studies is again significant and higher. The results of I^2 (Table 4) suggest that 32.862 % of the total variance can be attributed to the within-study sampling variance (level 1), 6.34 % to differences between effect sizes within studies (level 2) and 60.80 % to differences between studies (level 3). These results suggest that also for the long-term summary effect, the between-study heterogeneity is substantially higher than the within-study heterogeneity.

Table 4. *Long-term Pre-FU Summary Effect of Psilocybin- & MDMA- Assisted Therapy*

Model Results				
Hedges' *g*	*SE*	*t* (36)	*p*	*CI*
1.650	0.259	6.364	<.001***	1.117 - 2.183
Heterogeneity				
		Sigma	Sqrt	*p*
Within-study Variance (Level 2)		0.004	0.062	.4638
Between-study Variance (Level 3)		0.638	0.799	<.001***
Distribution of Variance (I^2)				
		I^2		
Amount Variance Level 1		32.862		
Amount Variance Level 2		6.340		
Amount Variance Level 3		60.798		

Note. Model Results display outputs of a three-level metanalytic model; *SE* = Standard Error; *CI* = Confidence Interval: heterogeneity was tested with one-sided log-likelihood-ratio tests; I^2 is a descriptive proportion that reflects variance in effects rather than sampling error; Significance codes: 0 '***'.

Moderator-analyses

Moderator-analyses were conducted on the short-term summary effect only because five primary studies were not included in the analysis of the long-term summary effect. The results of seven moderator-analyses are displayed in Table 5 and reported in the following.

3. Objective: Moderation of Acute Effects by 'Substance'

The omnibus *F*-test of the dichotomous moderator 'substance' (7 psilocybin studies; 9 MDMA studies) on the short-term summary effect is significant: $F(1,35) = 5.253$, $p = .028$, indicating that the changes in symptoms of mental disorders after psychedelic therapy are significantly moderated by the type of substance. This outcome supports the alternative hypothesis (H1: $p < .05$) to the third research question.

Contrasts of Moderator 'Substance'. The separate overall short-term summary effect of MDMA studies is Hedges' $g = 1.82$ and the separate short-term summary effect of psilocybin studies is Hedges' $g = 1.50$. The mean effect sizes of MDMA and psilocybin studies are significantly different from each other. $t(35) = 2.292$, $p = .028$. The difference in the short-term summary effects by substance per each effect size of each study can be viewed in the rain forest plot of the short-term summary effects (Figure 3).

4. Objective: Moderation of Acute Effects by 'Disorder'

The omnibus *F*-test of the categorical moderator 'disorder' (7 PTSD studies; 4 studies with anxiety and depression in patients with LTI; 3 studies with resistant MDD; 1 AUD study; 1 social anxiety study) on the short-term summary effect is significant: $F(4,32) = 5.614$, $p = .002$. This means that changes in symptoms of disorders after psychedelic therapy significantly differ by the type of disorder. This outcome supports the alternative hypothesis (H1: $p < .05$) based on the fourth research question.

Contrasts of Moderator 'Disorder'. The separate overall short-term summary effects of the disorders are: (1) resistant MDD $g = 2.082$, (2) anxiety and depression in patients with LTI $g = 1.398$, (3) PTSD $g = 1.741$, (4) AUD $g = 0.91$, and (5) social anxiety $g = 0.997$. These effects are all significantly different from each other, $F(4,32) = 5.614$, $p = .002$.

5. Objective: Moderation of Acute Effects by 'Study Design'

The omnibus *F*-test of the dichotomous moderator 'study design' (11 between-subjects studies; 5 within-subjects studies) is not significant: $F(1,35) = 1.622$, $p = .211$. This indicates that the changes in symptoms of mental disorders after psychedelic therapy are not moderated by the design of the clinical trial and that the acute summary effects of within-subject studies do not significantly differ from the acute summary effects of between-subject studies. This outcome supports the null hypothesis of the fifth research question (H0: $p \geq .05$).

6. Objective: Moderation of Acute Effects by 'Study Quality'

The omnibus *F*-test of the dichotomous moderator 'study quality' (11 low study quality; 5 high study quality) is not significant: $F(1,35) = 1.189$ $p = .283$. This means that the changes in symptoms of mental disorders after psychedelic therapy are not moderated by the categories of 'study quality' represented by the Cochrane's RoB tool (Higgins et al., 2011). The acute summary effects of studies with 'low study quality' ratings do not significantly differ from the acute summary effects of studies with 'high study quality' ratings. However, it is important to note that the interpretability of the rating of 'study quality' is limited because the Cochrane's RoB tool (Higgins et al., 2011) is designed for RCT studies only. This outcome supports the null hypothesis of the sixth research question (H0: $p \geq .05$).

Explorative Moderators

'Instrument'. The omnibus *F*-test of the explorative dichotomous moderator 'instrument' (30 self-report questionnaires; 7 clinician-administered interviews) is significant: $F(1,35) = 4.740$, $p = .036$. This indicates that the 30 endpoints measured by self-report questionnaires (BDI, BDI-II, STAI-S, STAI-T, AASE, PACS, QIDS-SR-16, IES-R, PDS) significantly differ from the seven endpoints measured by clinician-administered interviews (CAPS, CAPS-IV, CAPS-5). As the clinician-administered interviews are also based on the statements of the patients, it is important to note that the difference does not lie in the data source but in the administration of the assessment tool.

'Dosing'. The omnibus *F*-test of the explorative dichotomous moderator 'dosing' (3 moderate dosing studies; 13 high dosing studies) on the short-term summary effect is highly significant: $F(1, 35) = 23.501$, $p < .001$. This means that acute changes in symptoms of mental disorders in studies with moderate dosages of psychedelic substances only (psilocybin = 0.2 - 0.3 mg/kg; MDMA = 1.07 - 1.43 mg/kg) are significantly different from the acute summary effects in studies with high dosages of psychedelic substances (psilocybin = 0.3 - 0.43 mg/kg; MDMA = 1.43 mg/kg - 1.8 mg/kg). Only three primary studies used only moderate dosages of psychedelic substances. The other 13 studies that used moderate and high dosages of psychedelic substances were classified as 'high dosage' studies (see Table 1S).

'Peak experience effect'. Whether a peak experience has happened and correlated with the primary outcome of the therapeutic effect was only measured in studies administering psilocybin and not MDMA (see Table 1S). Therefore, the moderator 'peak experience effect' (6 assessments with no effect of peak experience on outcome; 12 assessments with effect of peak experience on outcome) was only tested on the seven psilocybin studies of the sample. The omnibus *F*-test of the explorative dichotomous moderator 'peak experience effect' on the

short-term summary effect is not significant: $F(1,15) = 3.688$, $p = .074$. This indicates that assessments that had a correlation effect with the primary outcome measures of the psilocybin studies did not significantly differ from the assessments that had no correlation effect with the primary outcome measures of psilocybin studies.

Table 5. *Moderation of the Pre-Post Summary Effect of Psilocybin- & MDMA-Therapy*

Moderator Variable	**Test of Moderators**			**Test of Residual Heterogeneity**		
	F	***df***	***p***	***Q***	***df***	***p***
Substance	5.253	1	.028*	120.500	35	< .001***
Disorder	5.614	4	.002**	103.296	32	< .001***
Study Design	1.622	1	.211	124.131	35	< .001***
Study Quality	1.189	1	.283	124.564	35	< .001***
Instrument	4.740	1	.036*	121.013	35	< .001***
Dosing	23.501	1	< .001***	102.252	35	< .001***
Peak Experience Effect	3.688	1	.074	50.635	15	< .001***

Note. F = estimate of omnibus-test of moderator added to the three-level model; df = degrees of freedom; Q = parameter of test of residual heterogeneity; Tests of residual heterogeneity show if there is significant unexplained variance left between all effect sizes after the moderator has been added to the model; Significance codes: 0 '***' 0.001 '**' 0.01 '*'.

Risk of Bias across Studies

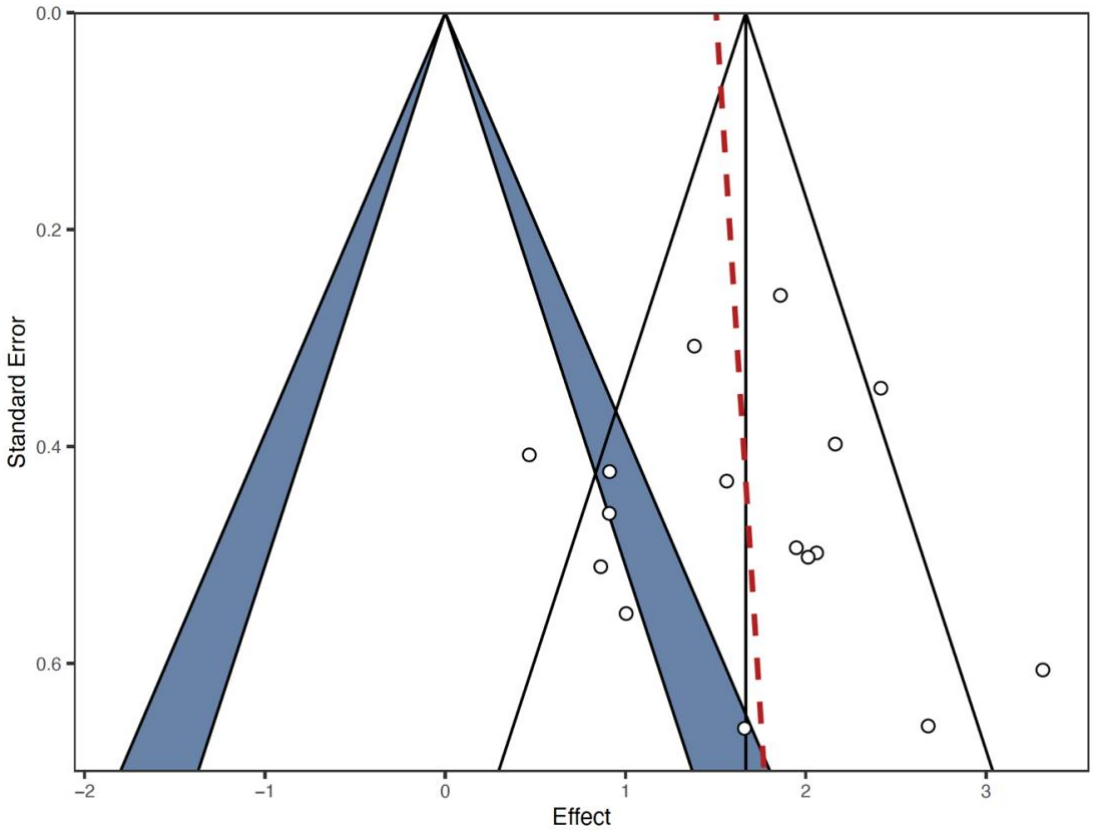
In the literature search and study selection process, efforts were made to control for a potential bias by looking for unpublished studies. No dissertations or theses of psychedelic clinical trials were found. Unpublished studies of the research registries were not provided and requests for data from unpublished MAPS studies were unanswered. It is therefore hard to say if there is a systematic bias in the reporting of large effects or significant results.

To conduct publication bias analyses, I aggregated all effect sizes of the short-term measures into one effect size per study to allow a more intuitive visual interpretation. The aggregation of effect sizes changed the short-term summary effects minimally in precision (Hedges'$g = 1.667$ instead of Hedges'$g = 1.684$ in the three-level model).

The funnel plot of the aggregated pre-post effects of all 16 primary studies is shown in Figure 5 and represents the relationship between the effect size and sample size (standard error) of each included study. In the case of a bias, more studies would be expected to the right and fewer to the left side towards the bottom of the plot because those on the left are more likely to be statistically insignificant and more likely to be missing (Borenstein et al., 2021). The pattern

of this funnel plot does therefore not resemble the pattern that is associated with bias. Supporting this visual interpretation, Egger's Regression test was not significant (t (14) = 0.243; p = .812), indicating no publication bias.

Figure 5. *Funnel Plot of the Aggregated Pre-Post Effects of the 16 Primary Studies*



Note. 'Effect' on the x-axis represents aggregated pre-post Hedges'*g* of the 16 primary studies. 'Standard error' on the y-axis represents the standard error of the aggregated effect size in each primary study with a reversed scale. Each of the 16 primary studies is represented by a dot. Large studies with low standard errors are shown at the top and small studies with high standard errors are shown at the bottom of the graph.

Discussion

In the following chapter, the results of the present meta-analysis will be summarized, discussed, and put into the context of established treatment effects for depression and PTSD. Ultimately, I will discuss the limitations of this thesis and give a brief outlook on the latest developments in the field of psychedelic research.

Summary of Evidence

Psilocybin and MDMA produce significant and very large effects one to six weeks after administration in a therapeutic setting. Symptom reduction of psychological disorders (PTSD, depression, anxiety, and AUD) persists over a long period of time - six to twelve months - with sustainably large and significant effects.

The type of 'substance' was a significant moderator of the acute summary effects and showed significantly higher effects in MDMA studies (g = 1.82) than in psilocybin studies (g = 1.50). Likewise, the type of 'disorder' treated was significantly moderating the acute summary effects with the highest treatment effects in patients with resistant MDD (g = 2.082) and lowest in patients with AUD (g = 0.91). Assumed moderators of relevance, 'study design' and 'study quality', were not found to be significant.

The explorative moderators 'instrument' and 'dosing' were significant. This indicates that the acute summary effect measured by the self-report questionnaires differ significantly from the effects measured by clinician-administered interviews, and that studies with moderate dosing have significantly different effects from studies that applied high dosing of psychedelic drugs. The explorative moderator 'peak experience effect' was not significant.

There was no indication of a publication bias across studies included, but nine out of 16 primary studies were conservatively rated as overall high potential risk of bias.

Discussion of Results

Large Effect Sizes

Considering the benchmarks for effect size interpretation by Cohen (1988), with .2, .5, and .8 counting as small, moderate, and large effects, respectively, the overall short-term Hedges' g = 1.684 and long-term Hedges' g = 1.650 can be regarded as remarkably high. Why are these treatment effects so high?

According to Borenstein et al. (2021), treatment effects are oftentimes inflated, when a new treatment is introduced, and samples of initial trials are relatively small. Out of safety

reasons, new interventions usually enroll a smaller number of patients that are very ill (see chapter Trial Phases). These patient groups show larger treatment benefits because they have more room to improve (Borenstein et al., 2021). Twelve of the 16 primary studies are those small-sampled (N = 10-30) treatment studies enrolling patients with treatment-resistant or terminal illnesses (see Table 2). This high proportion of small-sampled studies in the present meta-analysis could be a reason for the exceptionally high treatment effects.

Very large effect sizes could also be found in previous meta-analyses about psilocybin- and MDMA-assisted psychotherapy (Amoroso & Workman, 2016: Hedges' g = 1.17; Bahji et al., 2020: *SMD* = 1.30; Leudesdorff, 2021: Hedges' g = 1.17; Luoma et al., 2020: Hedges' g = 1.21; Romeo et al., 2020: *SMD* = 1.40). Within the same meta-analysis, comparisons within-subjects showed double as high treatment effects as between-subjects comparisons (Goldberg et al., 2020: within-subjects Hedges' g = 2.06 vs. between-subjects Hedges' g = 1.08). Hence, another potential reason for the large effect sizes could be the chosen within-subjects comparison.

Above all, it could also be possible that patients who enroll in psychedelic trials are psychedelic enthusiasts, believe in the effects of the therapy, and therefore show large symptom reduction (Ona et al., 2022). In any case, more large-sampled studies are necessary to draw meaningful conclusions.

Comparison of Psilocybin and MDMA Effect Sizes

Comparing the separate summary effects of psilocybin- and MDMA-studies (see Figures 3 and 4), the treatment effects of MDMA studies increased over time (from short-term Hedges' g = 1.82 to long-term g = 2.16), while the effects of psilocybin therapy dropped slightly after six to twelve months (from short-term Hedges' g = 1.50 to long-term g = 1.14). The long-term effects of MDMA studies are therefore almost double as high as the long-term effects of psilocybin studies. However, the confidence intervals of the estimates are larger in MDMA studies, and therefore the plausibility of 'true values' given the observed estimates is lower (see larger and more faded colors of raindrops in MDMA studies of Figures 3 and 4). Possible reasons for the difference in effect sizes of these two substances are yet to be addressed in the psychedelic research field.

Insignificant Within-study Heterogeneities

Both within-study heterogeneities are insignificant, indicating that the treatment outcomes between different instruments of the same study (level 2 of the three-level meta-analysis) are not significantly different from each other. According to Assink and Wibbelink (2016), this means that the fit of the three-level metanalytic model is not statistically better than

the fit of a conventional two-level model, and consequently, that the application of a three-level model might not be statistically necessary. This finding may vary depending on the inclusion of different instruments. The sensible application of the three-level meta-analysis to this data is ultimately up for discussion.

Possible Sources of Bias

Due to the small study effect, described by Sterne et al. (2001), the risk of publication bias increases with small-sampled studies. Larger studies usually invest more time and money and are therefore more likely to be published even if their results are negative (Sterne et al., 2001). That is why the funnel plot analyzes the relationship between effect size and sample size to detect chances of publication bias (see Figure 5).

Even though publication bias could not be found in this meta-analysis, the presence of other biases is quite realistic. Considering the risk of bias evaluations within each primary study (see Figure 2), the risk of a 'reporting bias' remains uncertain. Treatment outcomes of participants who dropped out of trials (see Table 1S, supplementary material) were often not included in outcomes analyses. This could have potentially confounded the results of separate primary studies.

Furthermore, the domain 'performance bias' was rated with the overall highest risk of bias (see Figure 2). Even though blinding took place in all primary studies with between-subjects design (11 out of 16 primary studies), the blinding was not successful in many of these studies (7 out of 11 blinded studies), meaning that participants and researchers were correctly guessing the treatment condition. If blinding was not successful in these studies, it means that patients knew that they had received the psychedelic substance. This can lead to a stronger rating of the symptom reduction and thus to an overestimation of the treatment effects. This risk is already inherent in the within-subjects design of the five open-label studies.

Effect Sizes in Context

In the following paragraphs, I will first briefly overview the treatment effects of established therapies for major depressive disorder and post-traumatic stress disorder. Then, I will briefly describe the few studies that have already compared psychedelic-assisted therapy to TAU.

Effect Sizes of Established Treatments for MDD and PTSD

Cognitive behavioral therapy (CBT) is the most thoroughly researched therapy for depression and has therefore a considerable weight of evidence (Cuijpers et al., 2013). In the meta-analysis of Cuijpers et al. (2013), the mean effect size of 75 CBT studies for depression

was Hedges' g = 0.71 (or g = 0.53 after adjustment for publication bias). The mean effect of this present meta-analysis (Hedges' g = 1.68) is more than double as high as this outcome.

To my knowledge, the largest network meta-analysis on the effects of common anti-depressant drugs analyzed 522 trials with 116477 MDD patients comparing the efficacy of 21 different SSRI drugs (Cipriani et al., 2018). The commonly administered SSRI antidepressants fluoxetine, citalopram, sertraline, and paroxetine showed small effects in patients with MDD compared to placebo (fluoxetine & citalopram: Cohen's d = 0.23; sertraline: d = 0.28; paroxetine: d = 0.31). The largest effect size was computed in the tricyclic antidepressant amitriptyline with an effect size of Cohen's d = 0.42 (Cipriani et al., 2018). All these effect sizes are four times smaller than the mean effect of this present meta-analysis (Hedges' g = 1.68).

A similar but smaller network meta-analysis was conducted by Cipriani et al. (2017) to assess the efficacy of SSRI drugs as medication for PTSD. The authors analyzed 51 clinical trials with 6189 participants comparing 25 different pharmacological treatments of PTSD. Again, only small symptom reductions were reported in PTSD patients after the administration of SSRIs fluoxetine, citalopram, sertraline, and paroxetine (sertraline: Cohen's d = 0.24; fluoxetine: d = 0.30; citalopram: d = 0.33; paroxetine: d = 0.38). The highest effect size of a drug on PTSD symptoms was computed in the non-selective and irreversible monoamine oxidase inhibitor phenelzine with an effect size of Cohen's d = 0.97 (Cipriani et al., 2017). The highest effect size is still almost half of the effect size that was achieved with MDMA-assisted psychotherapy.

However, the effect sizes of large-scale meta-analyses on established treatments of MDD and PTSD are hard to compare with the results of the present meta-analysis. Study designs and sample sizes vary drastically from psychedelic-assisted psychotherapy studies. Therefore, direct comparisons of psychedelic-assisted therapy with TAU can deliver more meaningful results.

Comparative Analyses of Psychedelic-Assisted Therapy with TAU

One primary study included in this present meta-analysis, the clinical trial by Carhart-Harris et al. (2021), compared the effects of psilocybin-assisted psychotherapy directly to the effects of the SSRI escitalopram on patients with MDD (N = 59). Symptom reduction occurred in 70% of the patients in the psilocybin group and in 48% of those in the escitalopram group, for a between-group difference of 22%. However, there was no significant difference between psilocybin and escitalopram treatments. According to the authors, larger and longer trials are necessary to compare psilocybin with conventional treatments (Carhart-Harris et al., 2021).

To my knowledge, no primary study has compared MDMA therapy effects directly to established psychopharmaceutics yet. However, the review of Feduccia et al., (2019) evaluated the efficacy of six MDMA 'phase 2' trials compared to two sertraline, and three paroxetine 'phase 3' trials. MDMA-assisted therapy showed a significant and large effect over the two established treatments (Cohen's $d = 0.9$; $p < .001$). The effect of MDMA was approximately double as big as paroxetine-effects and triple as big as sertraline-effects (Feduccia et al., 2019). In Addition to that, it is reported that MDMA-assisted therapy provided to patients with severe chronic PTSD is cost-saving from a payer's perspective (Marseille et al., 2022). Compared to the standard of care for 1000 patients, MDMA-assisted therapy allows net healthcare savings of $132.9 million over 30 years (Marseille et al., 2022).

Ultimately, more large-scale comparative analyses with established treatments are needed to make meaningful conclusions about the effectiveness of psilocybin- and MDMA-assisted therapy over TAU.

Limitations

Limitations of this Meta-Analysis

One methodological limitation of this present meta-analysis is that Cochrane's RoB (Higgins et al., 2011) was used for quality assessments for all included primary studies: RCTs and open-label studies (as in the meta-analysis by Goldberg et al., 2020). As a result of the conservative rating, all four open-label studies (Bogenschutz et al., 2015; Carhart-Harris et al., 2017; Monson et al., 2020; Wang et al., 2021) were rated with an overall high risk of bias. Random sequence generation, allocation concealment, and blinding of participants and researchers did not take place due to the open-label format (see Figure 2). Therefore, this evaluation lacks construct validity and the moderator 'study quality' lacks interpretability. Future quality assessments should consider differing study designs for more appropriate ratings.

Limitations of Psychedelic-Assisted Therapy Studies

The main methodological issues in the primary studies of this current meta-analysis are the already mentioned scarcity of studies and small sample sizes. It is therefore hard to make statistical inferences because the analysis might be underpowered, and the interpretability of results is accordingly limited. Furthermore, samples of psychedelic therapy studies might lack generalizability because approximately only a fraction of subjects are psychedelic-naïve (Ona et al., 2022). In most primary studies of this meta-analysis, participants either had previous experience with psychedelic drugs or consider their mind-altering effects desirable. Another important limitation is that most included studies applied per-protocol instead of intention-to-

treat analysis. This can be dangerous because it potentially leads to an overrating of the drug's safety profile and an underreporting of adverse events (Ona et al., 2022).

Considering its history of rigorous stigmatization, it is important to provide a high standard of scientific rigor and transparency in psychedelic research (Petranker et al., 2020). According to their pragmatic research checklist, transparency standards of Open Science using preregistration, open materials, and data are highly recommended. Apart from that, constraints on generalizability should always be reported and replication of results should always be encouraged (Petranker et al., 2020).

Latest Developments

The latest news from the Multidisciplinary Association for Psychedelic Studies reveals that prior significant and large effects of the first-ever 'phase 3' MDMA clinical trial by Mitchell et al. (2021) were successfully confirmed in a second 'phase 3' MDMA clinical trial. 104 PTSD patients were either treated with MDMA-assisted therapy or placebo therapy. This represents another important milestone. Expected to be published later this year, these findings will support the FDA's new drug application of MDMA (Aldworth, 2023).

Conclusion

Psilocybin and MDMA produce significant and very large effects in treatment settings that persist over six to twelve months. More clinical trials with larger samples and direct comparisons with established treatments are needed to make meaningful inferences about psychedelic treatments. Even though the interpretability of results is still limited due to studies with mostly small and selective samples, psilocybin and especially MDMA studies show great potential for prospective psychopharmacological solutions. Considering the trend of ongoing crises worldwide (pandemics, climate crises, and wars), new perspectives on effective psychological support are urgently needed. In current times of emotional and physical disconnection, the holistic approach of psychedelic treatment might contribute to a greater sense of belonging and healing.

References

Aday, J. S., Mitzkovitz, C. M., Bloesch, E. K., Davoli, C. C., & Davis, A. K. (2020). Long-term effects of psychedelic drugs: A systematic review. *Neuroscience & Biobehavioral Reviews*, *113*, 179–189. https://doi.org/10.1016/j.neubiorev.2020.03.017

Aldworth, B. (2023, January 7). *Prior positive results confirmed in MAPS-sponsored, philanthropy-funded phase 3 trial*. Multidisciplinary Association for Psychedelic Studies - MAPS. https://maps.org/2023/01/05/prior-positive-results-confirmed/

Amoroso, T., & Workman, M. (2016). Treating posttraumatic stress disorder with MDMA-assisted psychotherapy: A preliminary meta-analysis and comparison to prolonged exposure therapy. *Journal of Psychopharmacology*, *30*(7), 595–600. https://doi.org/10.1177/0269881116642542

Andersen, K. K., Carhart-Harris, R. L., Nutt, D. J., & Erritzoe, D. (2021). Therapeutic effects of classic serotonergic psychedelics: A systematic review of modern-era clinical studies. *Acta Psychiatrica Scandinavica*, *143*(2), 101–118. https://doi.org/10.1111/acps.13249

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, *1316*(1), 29–52. https://doi.org/10.1111/nyas.12360

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

Bahji, A., Forsyth, A., Groll, D., & Hawken, E. R. (2020). Efficacy of 3,4-methylenedioxymethamphetamine (MDMA)-assisted psychotherapy for posttraumatic stress disorder: A systematic review and meta-analysis. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *96*, 109735. https://doi.org/10.1016/j.pnpbp.2019.109735

Barrowman, N. J., & Myers, R. A. (2003). Raindrop plots. *The American Statistician*, *57*(4), 268–274. https://doi.org/10.1198/0003130032369

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

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbauch, J. (1961). *Beck Depression Inventory (BDI)* [Database record]. APA PsycTests. https://doi.org/10.1037/t00741-000

Bedi, G., Hyman, D. M., & De Wit, H. (2010). Is ecstasy an "empathogen"? Effects of ±3,4-methylenedioxymethamphetamine on prosocial feelings and identification of emotional states in others. *Biological Psychiatry*, *68*(12), 1134–1140. https://doi.org/10.1016/j.biopsych.2010.08.003

Benzenhöfer, U., & Passie, T. (2006). Zur Frühgeschichte von „Ecstasy". *Der Nervenarzt*, *77*(1), 95–99. https://doi.org/10.1007/s00115-005-2001-y

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress*, *8*(1), 75-90. https://doi.org/10.1002/jts.2490080106

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2nd ed.). Wiley.

Card, N. A. (2012). *Applied meta-analysis for social science research.* Guilford Press.

Carhart-Harris, R. L., Bolstridge, M., Day, C. M. J., Rucker, J., Watts, R., Erritzoe, D. E., Kaelen, M., Giribaldi, B., Bloomfield, M., Pilling, S., Rickard, J. A., Forbes, B., Feilding, A., Taylor, D., Curran, H. V., & Nutt, D. J. (2017). Psilocybin with psychological support for treatment-resistant depression: Six-month follow-up.

Psychopharmacology, *235*(2), 399–408. https://doi.org/10.1007/s00213-017-4771-x

Carhart-Harris, R. L., Bolstridge, M., Rucker, J., Day, C. M., Erritzoe, D., Kaelen, M., Bloomfield, M., Rickard, J. A., Forbes, B., Feilding, A., Taylor, D., Pilling, S., Curran, V., & Nutt, D. J. (2016). Psilocybin with psychological support for treatment-resistant depression: An open-label feasibility study. *The Lancet Psychiatry*, *3*(7), 619–627. https://doi.org/10.1016/s2215-0366(16)30065-7

Carhart-Harris, R., Giribaldi, B., Watts, R., Baker-Jones, M., Murphy-Beiner, A., Murphy, R., Martell, J., Blemings, A., Erritzoe, D., & Nutt, D. J. (2021). Trial of psilocybin versus escitalopram for depression. *New England Journal of Medicine*, *384*(15), 1402–1411. https://doi.org/10.1056/nejmoa2032994

Carhart-Harris, R. L., & Goodwin, G. M. (2017). The therapeutic potential of psychedelic drugs: Past, present, and future. *Neuropsychopharmacology*, *42*(11), 2105–2113. https://doi.org/10.1038/npp.2017.84

Carhart-Harris, R. L., Murphy, K., Leech, R., Erritzoe, D., Wall, M. B., Ferguson, B., Williams, L., Roseman, L., Brugger, S., De Meer, I., Tanner, M. A., Tyacke, R. J., Wolff, K., Sethi, A., Bloomfield, M., Williams, T., Bolstridge, M., Stewart, L., Morgan, C. J. A., . . . Nutt, D. J. (2015). The effects of acutely administered 3,4-methylenedioxymethamphetamine on spontaneous brain function in healthy volunteers measured with arterial spin labeling and blood oxygen level–dependent resting state functional connectivity. *Biological Psychiatry*, *78*(8), 554–562. https://doi.org/10.1016/j.biopsych.2013.12.015

Carod-Artal, F. J. (2015). Hallucinogenic drugs in pre-Columbian Mesoamerican cultures. *Neurologia*, *30*(1), 42–49. https://doi.org/10.1016/j.nrl.2011.07.003

Cheung, M. W. L. (2015). *Meta-analysis: A structural equation modeling approach*. John Wiley & Sons.

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., Leucht, S., Ruhé, H. G., Turner, E. H., Higgins, J. P. T., Egger, M., Takeshima, N., Hayasaka, Y., Imai, H., Shinohara, K., Tajika, A., Ioannidis, J. P. A., & Geddes, J. R. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute

treatment of adults with major depressive disorder: A systematic review and network meta-analysis. *The Lancet*, *391*(10128), 1357–1366. https://doi.org/10.1016/s0140-6736(17)32802-7

Cipriani, A., Williams, T., Nikolakopoulou, A., Salanti, G., Chaimani, A., Ipser, J. C., Cowen, P. J., Geddes, J. R., & Stein, D. J. (2017). Comparative efficacy and acceptability of pharmacological treatments for post-traumatic stress disorder in adults: A network meta-analysis. *Psychological Medicine*, *48*(12), 1975–1984. https://doi.org/10.1017/s003329171700349x

Clarivate. (2021). *Endnote X9* (Version 3.1). [Computer software]. https://endnote.com/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Comprehensive R Archive Network (CRAN). (2020). *CRAN - package metaviz*. [Computer Software]. https://cran.r-project.org/web/packages/metaviz/index.html

Cooper, H., Hedges, L.V., & Valentine, J.C. (2019). *The handbook of research synthesis and meta-analysis*. (3rd ed.). Russell Sage Foundation.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry*, *58*(7), 376–385. https://doi.org/10.1177/070674371305800702

Danforth, A. L., Grob, C. S., Struble, C., Feduccia, A. A., Walker, N., Jerome, L., Yazar-Klosinski, B., & Emerson, A. (2018). Reduction in social anxiety after MDMA-assisted psychotherapy with autistic adults: A randomized, double-blind, placebo-controlled pilot study. *Psychopharmacology*, *235*(11), 3137–3148. https://doi.org/10.1007/s00213-018-5010-9

Davis, A. K., Barrett, F. S., May, D. G., Cosimano, M. P., Sepeda, N. D., Johnson, M. W., Finan, P. H., & Griffiths, R. R. (2021). Effects of psilocybin-assisted therapy on major depressive disorder. *JAMA Psychiatry*, *78*(5), 481. https://doi.org/10.1001/jamapsychiatry.2020.3285

Daws, R. E., Timmermann, C., Giribaldi, B., Sexton, J., Wall, M. B., Erritzoe, D., Roseman, L., Nutt, D. J., & Carhart-Harris, R. L. (2022). Increased global integration in the brain after psilocybin therapy for depression. *Nature Medicine*, *28*(4), 844–851. https://doi.org/10.1038/s41591-022-01744-z

DeAngelis, T. (2019, November 1). *Better relationships with patients lead to better outcomes*. https://www.apa.org/monitor/2019/11/ce-corner-relationships

DiClemente, C. C., Carbonari, J. P., Montgomery, R. P. G., & Hughes, S. O. (1994). The Alcohol Abstinence Self-Efficacy Scale. *Journal of Studies on Alcohol*, *55*(2), 141–148. https://doi.org/10.15288/jsa.1994.55.141

Doblin, R. E., Christiansen, M., Jerome, L., & Burge, B. (2019). The past and future of psychedelic science: An introduction to this issue. *Journal of Psychoactive Drugs*, *51*(2), 93–97. https://doi.org/10.1080/02791072.2019.1606472

Dos Santos, R. G., & Hallak, J. E. C. (2020). Therapeutic use of serotoninergic hallucinogens: A review of the evidence and of the biological and psychological mechanisms. *Neuroscience & Biobehavioral Reviews*, *108*, 423–434. https://doi.org/10.1016/j.neubiorev.2019.12.001

Dos Santos, R. G., Osório, F. L., Crippa, J. A. S., Riba, J., Zuardi, A. W., & Hallak, J. E. C. (2016). Antidepressive, anxiolytic, and antiaddictive effects of ayahuasca, psilocybin and lysergic acid diethylamide (LSD): A systematic review of clinical trials published in the last 25 years. *Therapeutic Advances in Psychopharmacology*, *6*(3), 193–213. https://doi.org/10.1177/2045125316638008

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Euronews. (2022, October 10). *Mental health day: How war in Ukraine is causing psychological trauma for millions of people.* https://www.euronews.com/2022/10/10/mental-health-day-how-war-in-ukraine-is causing-psychological-trauma-for-millions-of-peopl

Feduccia, A. A., Jerome, L., Yazar-Klosinski, B., Emerson, A., Mithoefer, M. C., & Doblin, R. (2019). Breakthrough for trauma treatment: Safety and efficacy of MDMA-assisted psychotherapy compared to paroxetine and sertraline. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00650

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). Visual representations of meta-analyses of multiple outcomes: Extensions to forest plots, funnel plots, and caterpillar plots. *Methodology*, *16*(4), 299–315. https://doi.org/10.5964/meth.4013

Flannery, B., Volpicelli, J. R., & Pettinati, H. M. (1999). Psychometric properties of the Penn Alcohol Craving Scale. *Alcoholism: Clinical and Experimental Research*, *23*(8), 1289. https://doi.org/10.1097/00000374-199908000-00001

Foa, E. B., Cashman, L., Jaycox, L. H., & Perry, K. I. (1997). The validation of a self-report measure of posttraumatic stress disorder: The Posttraumatic Diagnostic Scale. *Psychological Assessment*, *9*(4), 445–451. https://doi.org/10.1037/1040-3590.9.4.445

Foa, E., Hembree, E., & Rothbaum, B. O. (2007). *Prolonged exposure therapy for PTSD: Emotional processing of traumatic experiences*. Oxford University Press.

Gable, R. S. (1993). Toward a comparative overview of dependence potential and acute toxicity of psychoactive substances used nonmedically. *The American Journal of Drug and Alcohol Abuse*, *19*(3), 263–281. https://doi.org/10.3109/00952999309001618

Galvão-Coelho, N. L., Marx, W., Gonzalez, M. C., Sinclair, J., De Manincor, M. J., Perkins, D. F., & Sarris, J. (2021). Classic serotonergic psychedelics for mood and depressive symptoms: A meta-analysis of mood disorder patients and healthy participants. *Psychopharmacology*, *238*(2), 341–354. https://doi.org/10.1007/s00213-020-05719-1

Garcia-Romeu, A., & Richards, W. A. (2018). Current perspectives on psychedelic therapy: Use of serotonergic hallucinogens in clinical interventions. *International Review of Psychiatry*, *30*(4), 291–316. https://doi.org/10.1080/09540261.2018.1486289

Goldberg, S. B., Shechet, B., Nicholas, C. R., Ng, C. W., Deole, G., Chen, Z., & Raison, C. L. (2020). Post-acute psychological effects of classical serotonergic psychedelics: A systematic review and meta-analysis. *Psychological Medicine*, *50*(16), 2655–2666. https://doi.org/10.1017/S003329172000389X

Greer, G., & Tolbert, R. (1998). A method of conducting therapeutic sessions with MDMA. *Journal of Psychoactive Drugs*, *30*(4), 371–379. https://doi.org/10.1080/02791072.1998.10399713

Griffiths, R. R., Johnson, M. W., Carducci, M. A., Umbricht, A., Richards, W. A., Richards, B. D., Cosimano, M. P., & Klinedinst, M. A. (2016). Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: A randomized double-blind trial. *Journal of Psychopharmacology*, *30*(12), 1181–1197. https://doi.org/10.1177/0269881116675513

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

Guina, J., Rossetter, S. R., De Rhodes, B. J., Nahhas, R. W., & Welton, R. S. (2015). Benzodiazepines for PTSD. *Journal of Psychiatric Practice*, *21*(4), 281–303. https://doi.org/10.1097/pra.0000000000000091

Halberstadt, A. L. (2015). Recent advances in the neuropsychopharmacology of serotonergic hallucinogens. *Behavioural Brain Research*, *277*(1), 99–120. https://doi.org/10.1016/j.bbr.2014.07.016

Halpern, J. H., & Pope, H. G. (2003). Hallucinogen persisting perception disorder: What do we know after 50 years? *Drug and Alcohol Dependence*, *69*(2), 109–119. https://doi.org/10.1016/s0376-8716(02)00306-x

Hamilton, J. H., Furman, D. J., Chang, C., Thomason, M. E., Dennis, E. L., & Gotlib, I. H. (2011). Default-mode and task-positive network activity in major depressive disorder: Implications for adaptive and maladaptive rumination. *Biological Psychiatry*, *70*(4), 327–333. https://doi.org/10.1016/j.biopsych.2011.02.003

Hartogsohn, I. (2016). Set and setting, psychedelics and the placebo response: An extra-pharmacological perspective on psychopharmacology. *Journal of Psychopharmacology*, *30*(12), 1259–1267. https://doi.org/10.1177/0269881116677852

Hartogsohn, I. (2017). Constructing drug effects: A history of set and setting. *Drug Science, Policy and Law*, *3*. https://doi.org/10.1177/2050324516683325

Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., Savovic, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*(oct18 2), 5928–5928. https://doi.org/10.1136/bmj.d5928

Hofmann, A. F., Heim, R., Brack, A., & Kobel, H. R. (1958). Psilocybin, ein psychotroper Wirkstoff aus dem mexikanischen Rauschpilz Psilocybe mexicana Heim. *Experientia*, *14*(3), 107–109. https://doi.org/10.1007/bf02159243

Hoskins, M., Sinnerton, R., Nakamura, A., Underwood, J., Slater, A., Lewis, C., Roberts, N., Bisson, J. I., Lee, M. K. O., & Clarke, L. (2021). Pharmacological-assisted psychotherapy for post-traumatic stress disorder: A systematic review and meta-analysis. *European Journal of Psychotraumatology*, *12*(1). https://doi.org/10.1080/20008198.2020.1853379

Illingworth, B. J., Lewis, D. J., Lambarth, A., Stocking, K., Duffy, J., Jelen, L. A., & Rucker, J. (2021). A comparison of MDMA-assisted psychotherapy to non-assisted psychotherapy in treatment-resistant PTSD: A systematic review and meta-analysis. *Journal of Psychopharmacology*, *35*(5), 501–511. https://doi.org/10.1177/0269881120965915

Ioannidis J. P. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly*, *94*(3), 485–514. https://doi.org/10.1111/1468-0009.12210

Jerome, L., Feduccia, A. A., Wang, J., Hamilton, S., Yazar-Klosinski, B., Emerson, A., Mithoefer, M. C., & Doblin, R. (2020). Long-term follow-up outcomes of MDMA-assisted psychotherapy for treatment of PTSD: A longitudinal pooled analysis of six phase 2 trials. *Psychopharmacology*, *237*(8), 2485–2497.

https://doi.org/10.1007/s00213-020-05548-2

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

Johnson, M. P., Griffiths, R. R., Hendricks, P. S., & Henningfield, J. E. (2018). The abuse potential of medical psilocybin according to the 8 factors of the Controlled Substances Act. *Neuropharmacology*, *142*, 143–166. https://doi.org/10.1016/j.neuropharm.2018.05.012

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

Jonas, D. E., Cusack, K., Forneris, C. A., Wilkins, T. M., Sonis, J., Middleton, J. C., Feltner, C., Meredith, D., Cavanaugh, J., Brownley, K. A., Olmsted, K. R., Greenblatt, A., Weil, A., & Gaynes, B. N. (2013). *Psychological and pharmacological treatments for adults with posttraumatic stress disorder (PTSD)*. Agency for Healthcare Research and Quality (US). https://www.ncbi.nlm.nih.gov/books/NBK137702/pdf/Bookshelf_NBK137702.pdf

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Kolovos, S., Van Tulder, M. W., Cuijpers, P., Prigent, A., Chevreul, K., Riper, H., & Bosmans, J. E. (2017). The effect of treatment as usual on major depressive disorder: A meta-analysis. *Journal of Affective Disorders*, *210*, 72–81. https://doi.org/10.1016/j.jad.2016.12.013

Leudesdorff, M. T. (2021). *Meta-analysis on the effects of psychedelics- and MDMA-assisted therapy in clinical settings*. [Unpublished master thesis]. Universität Wien.

Liechti, M. E., Baumann, C., Gamma, A., & Vollenweider, F. X. (2000). Acute psychological effects of 3,4-methylenedioxymethamphetamine (MDMA, “Ecstasy”) are attenuated by the serotonin uptake inhibitor citalopram. *Neuropsychopharmacology*, *22*(5),

513–521. https://doi.org/10.1016/s0893-133x(99)00148-7

Luoma, J. B., Chwyl, C., Bathje, G. J., Davis, A. T., & Lancelotta, R. (2020). A meta-analysis of placebo-controlled trials of psychedelic-assisted therapy. *Journal of Psychoactive Drugs*, *52*(4), 289–299. https://doi.org/10.1080/02791072.2020.1769878

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

MAPS. (2022, March 30). *Benzinga: World's first phase 3 clinical trial on MDMA shows "powerful new pathway to healing".* https://maps.org/news/media/benzinga-world-s-first-phase-3-clinical-trial-on-mdma-shows-powerful-new-pathway-to-healing/

Marseille, E., Mitchell, J. P., & Kahn, J. G. (2022). Updated cost-effectiveness of MDMA-assisted therapy for the treatment of posttraumatic stress disorder in the United States: Findings from a phase 3 trial. *PLOS ONE*, *17*(2), e0263252. https://doi.org/10.1371/journal.pone.0263252

Maslow, A. H. (1959). Cognition of being in the peak experiences. *Journal of Genetic Psychology*, *94*(1), 43–66. https://doi.org/10.1080/00221325.1959.10532434

McKenzie, J. E., Brennan, S. E., Ryan, R. E., Thomson, H. J., Johnston, R. V., Thomas, J. (2021). Defining the criteria for including studies and how they will be grouped for the synthesis. In Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., Welch, V.A. (Eds.), *Cochrane handbook for systematic reviews of interventions* (Version 6.2). Cochrane. www.training.cochrane.org/handbook.

Mitchell, J. M., Bogenschutz, M., Lilienstein, A., Harrison, C., Kleiman, S., Parker-Guilbert, K., Ot'alora G., M., Garas, W., Paleos, C., Gorman, I., Nicholas, C., Mithoefer, M., Carlin, S., Poulter, B., Mithoefer, A., Quevedo, S., Wells, G., Klaire, S. S., van der Kolk, B., . . . Doblin, R. (2021). MDMA-assisted therapy for severe PTSD: A randomized, double-blind, placebo-controlled phase 3 study. *Nature Medicine*, *27*(6), 1025–1033. https://doi.org/10.1038/s41591-021-01336-3

Mithoefer, M. C., Mithoefer, A. T., Feduccia, A. A., Jerome, L., Wagner, M., Wymer, J., Holland, J., Hamilton, S., Yazar-Klosinski, B., Emerson, A., & Doblin, R. (2018). 3,4-methylenedioxymethamphetamine (MDMA)-assisted psychotherapy for post-traumatic stress disorder in military veterans, firefighters, and police officers: A randomised, double-blind, dose-response, phase 2 clinical trial. *The Lancet Psychiatry*, *5*(6), 486–497. https://doi.org/10.1016/s2215-0366(18)30135-4

Mithoefer, C., Mithoefer, M. D., & Multidisciplinary Association for Psychedelic Studies. (2017). *A manual for MDMA-assisted psychotherapy in the treatment of posttraumatic stress disorder*. https://s3-us-west-1.amazonaws.com/mapscontent/research-archive/mdma/TreatmentManual_MDMAAssistedPsychotherapyVersion+8.1_22+Aug2017.pdf

Mithoefer, M. C., Wagner, M. T., Mithoefer, A. T., Jerome, L., & Doblin, R. (2010). The safety and efficacy of ±3,4-methylenedioxymethamphetamine-assisted psychotherapy in subjects with chronic, treatment-resistant posttraumatic stress disorder: The first randomized controlled pilot study. *Journal of Psychopharmacology*, *25*(4), 439–452. https://doi.org/10.1177/0269881110378371

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Monson, C. M., Wagner, A. C., Mithoefer, A. T., Liebman, R. E., Feduccia, A. A., Jerome, L., Yazar-Klosinski, B., Emerson, A., Doblin, R., & Mithoefer, M. C. (2020). MDMA-facilitated cognitive-behavioural conjoint therapy for posttraumatic stress disorder: An uncontrolled trial. *European Journal of Psychotraumatology*, *11*(1), 1840123. https://doi.org/10.1080/20008198.2020.1840123

Murphy, R., Kettner, H., Zeifman, R., Giribaldi, B., Kartner, L., Martell, J., Read, T. R. H., Murphy-Beiner, A., Baker-Jones, M., Nutt, D. J., Erritzoe, D., Watts, R., & Carhart-Harris, R. L. (2022). Therapeutic alliance and rapport modulate responses to psilocybin assisted therapy for depression. *Frontiers in Pharmacology*, *12*. https://doi.org/10.3389/fphar.2021.788155

Nardou, R., Lewis, E. M., Rothhaas, R., Xu, R., Yang, A., Boyden, E. S., & Dölen, G. (2019). Oxytocin-dependent reopening of a social reward learning critical period with MDMA. *Nature*, *569*(7754), 116–120. https://doi.org/10.1038/s41586-019-1075-9

National Institute on Drug Abuse (NIDA). (2017, February 23). *MDMA (Ecstasy) Abuse Research Report.* National Institute on Drug Abuse. https://nida.nih.gov/download/1763/mdma-ecstasy-abuse-research-report.pdf?v=841ad18273f37a278f164eb491d196db

Nichols, D. E. (2016). Psychedelics. *Pharmacological Reviews*, *68*(2), 264–355. https://doi.org/10.1124/pr.115.011478

Nutt, D. (2021). *Drugs without the hot air: Minimising the harms of legal and illegal drugs*. Cambridge University Press.

Nutt, D. J., King, L. A., & Phillips, L. S. (2010). Drug harms in the UK: A multicriteria decision analysis. *The Lancet*, *376*(9752), 1558–1565. https://doi.org/10.1016/s0140-6736(10)61462-6

Oehen, P., Traber, R., Widmer, V., & Schnyder, U. (2013). A randomized, controlled pilot study of MDMA (±3,4-Methylenedioxymethamphetamine)-assisted psychotherapy for treatment of resistant, chronic post-traumatic stress disorder (PTSD). *Journal of Psychopharmacology*, *27*(1), 40–52. https://doi.org/10.1177/0269881112464827

Ona, G., Kohek, M., & Bouso, J. C. (2022). The illusion of knowledge in the emerging field of psychedelic research. *New Ideas in Psychology*, *67*, 100967. https://doi.org/10.1016/j.newideapsych.2022.100967

Orsolini, L., Papanti, G. D., De Berardis, D., Guirguis, A., Corkery, J. M., & Schifano, F. (2017). The “endless trip” among the NPS users: Psychopathology and psychopharmacology in the hallucinogen-persisting perception disorder. A systematic review. *Frontiers in Psychiatry*, *8*. https://doi.org/10.3389/fpsyt.2017.00240

Ot’alora G, M., Grigsby, J., Poulter, B., Van Derveer, J. W., Giron, S. G., Jerome, L., Feduccia, A. A., Hamilton, S., Yazar-Klosinski, B., Emerson, A., Mithoefer, M. C., & Doblin, R. (2018). 3,4-Methylenedioxymethamphetamine-assisted psychotherapy for

treatment of chronic posttraumatic stress disorder: A randomized phase 2 controlled trial. *Journal of Psychopharmacology*, *32*(12), 1295–1307. https://doi.org/10.1177/0269881118806297

Ott, J., & Bigwood, J. (1978). *Teonanácatl: Hallucinogenic mushrooms of North America*. Madrona Publishers.

Ottawa Hospital Research Institute. (2011). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses.* https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 71. https://doi.org/10.1136/bmj.n71

Papaseit, E., Torrens, M., Pérez-Mañá, C., Muga, R., & Farré, M. (2018). Key interindividual determinants in MDMA pharmacodynamics. *Expert Opinion on Drug Metabolism & Toxicology*, *14*(2), 183–195. https://doi.org/10.1080/17425255.2018.1424832

Petranker, R., Anderson, T., & Farb, N. (2020). Psychedelic research and the need for transparency: Polishing Alice's looking glass. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01681

Pollan, M. (2019). *How to change your mind: What the new science of psychedelics teaches us about consciousness, dying, addiction, depression, and transcendence*. Penguin Books.

R Development Core Team. (2020). *R: A language and environment for statistical computing* (Version 3.6.3). [Computer Software]. http://www.R-project.org/

Rätsch, C. (2018). *Enzyklopädie der psychoaktiven Pflanzen: Botanik, Ethnopharmakologie und Anwendung*. AT Verlag.

Reiff, C. M., Richman, E. E., Nemeroff, C. B., Carpenter, L. L., Widge, A. S., Rodriguez, C. I., Kalin, N. H., & McDonald, W. M. (2020). Psychedelics and psychedelic-assisted psychotherapy. *American Journal of Psychiatry*, *177*(5), 391–410.

https://doi.org/10.1176/appi.ajp.2019.19010035

Risk of bias tools. (n.d.). *Robvis (visualization tool)*. [Computer Software]. https://www.riskofbias.info/welcome/robvis-visualization-tool

Rohatgi, A. (2020). *WebPlotDigitizer* (Version 4.6). [Computer Software]. https://automeris.io/WebPlotDigitizer/

Romeo, B., Karila, L., Martelli, C., & Benyamina, A. (2020). Efficacy of psychedelic treatments on depressive symptoms: A meta-analysis. *Journal of Psychopharmacology*, *34*(10), 1079–1085. https://doi.org/10.1177/0269881120919957

Roseman, L., Nutt, D. J., & Carhart-Harris, R. L. (2018). Quality of acute psychedelic experience predicts therapeutic efficacy of psilocybin for treatment-resistant depression. *Frontiers in Pharmacology*, *8*. https://doi.org/10.3389/fphar.2017.00974

Ross, S., Bossis, A., Guss, J., Agin-Liebes, G., Malone, T., Cohen, B., Mennenga, S. E., Belser, A., Kalliontzi, K., Babb, J., Su, Z., Corby, P., & Schmidt, B. L. (2016). Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: A randomized controlled trial. *Journal of Psychopharmacology*, *30*(12), 1165–1180. https://doi.org/10.1177/0269881116675512

Rucker, J. J., Iliff, J., & Nutt, D. J. (2018). Psychiatry & the psychedelic drugs. Past, present & future. *Neuropharmacology*, *142*, 200–218. https://doi.org/10.1016/j.neuropharm.2017.12.040

Rucker, J., Schnall, J., D'Hotman, D., King, D., Davis, T., & Neill, J., (2020). *Medicinal use of psilocybin: Reducing restrictions on research and treatment.* Adam Smith Research Trust. https://static1.squarespace.com/static/5bbb29273560c345fcc0fade/t/5f32545b5127cf236590f0bb/1597133926474/Medicinal+Use+of+Psilocybin+-+Rucker%2C+Schnall%2C+D%27Hotman%2C+King%2C+Davis+and+Neill.pdf

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B. A., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S. G., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-Item Quick Inventory of Depressive

Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, *54*(5), 573–583. https://doi.org/10.1016/s0006-3223(02)01866-8

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory.* Consulting Psychologists Press.

Sterne, J. A. C., Egger, M., & Smith, G. D. (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ*, *323*(7304), 101–105. https://doi.org/10.1136/bmj.323.7304.101

Studerus, E., Kometer, M., Hasler, F., & Vollenweider, F. X. (2010). Acute, subacute and long-term subjective effects of psilocybin in healthy humans: A pooled analysis of experimental studies. *Journal of Psychopharmacology*, *25*(11), 1434–1452. https://doi.org/10.1177/0269881110382466

Tedesco, S., Gajaram, G., Chida, S., Ahmad, A., Pentak, M., Kelada, M., Lewis, L., Krishnan, D., Tran, C., Soetan, O. T., Mukona, L. T., & Jolayemi, A. (2021). The efficacy of MDMA (3,4-Methylenedioxymethamphetamine) for post-traumatic stress disorder in humans: A systematic review and meta-analysis. *Cureus*, *13*(5), e15070. https://doi.org/10.7759/cureus.15070

Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., Norquist, G., Howland, R. H., Lebowitz, B. D., McGrath, P. J., Shores-Wilson, K., Biggs, M. M., Balasubramani, G. K., & Fava, M. (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice. *American Journal of Psychiatry*, *163*(1), 28–40. https://doi.org/10.1176/appi.ajp.163.1.28

U.S. Food and Drug Administration. (2018, January 4). *Step 3: Clinical Research*. https://www.fda.gov/patients/drug-development-process/step-3-clinical-research

Van Den Noortgate, W., López-López, J., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, *47*(4), 1274–1294. https://doi.org/10.3758/s13428-014-0527-2

Viechtbauer, W. (2015). *Meta-analysis package for R – metafor*. [Computer Software]. https://cran.rproject.org/web/packages/metafor/metafor.pdf

Wang, J. B., Lin, J., Bedrosian, L., Coker, A., Jerome, I., Feduccia, A., Lilienstein, A., Harrison, C., Heimler, E., Mithoefer, M., Mithoefer, A., Ot'alora G., M., Poulter, B., Carlin, S., Matthews, R., Yazar-Klosinski, B., Emerson, A., & Doblin, R. (2021). Scaling up: Multisite open-label clinical trials of MDMA-assisted therapy for severe posttraumatic stress disorder. *Journal of Humanistic Psychology, 0*(0), 1-29. https://doi.org/10.1177/00221678211023663

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013). *The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)*. [Assessment] Available from www.ptsd.va.gov.

Weiss, D. S., & Marmar, C. R. (1996). The Impact of Event Scale-Revised. In Wilson, J. & Keane, T. M. (Eds.), *Assessing psychological trauma and PTSD*. Guilford.

Wolfson, P. E., Andries, J., Feduccia, A. A., Jerome, L., Wang, J. B., Williams, E., Carlin, S. C., Sola, E., Hamilton, S., Yazar-Klosinski, B., Emerson, A., Mithoefer, M. C., & Doblin, R. (2020). MDMA-assisted psychotherapy for treatment of anxiety and other psychological distress related to life-threatening illnesses: A randomized pilot study. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-75706-1

World Health Organization (WHO). (2022, March 2). *COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide*. https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

Supplementary Material

Table 1S. *Study Characteristics of 16 Primary Studies Included in Three-Level Meta-Analysis*

Study	Dosing	Course of Trial	Available Data	Instruments	Assessment Time Points	Peak Experience	Dropouts & Adverse Effects
Grob et al. (2011) psilocybin anxiety & depression in cancer patients *N* = 12 RCT, but within subjects	moderate dose (0.2 mg/kg)	2 dosing sessions several weeks apart no PT same subject gets 1x psilocybin and 1x placebo randomization of order which substance first	no outcomes in tables of study email sent to Grob via researchgate.net data won with **WebPlot Digitizer**	BDI STAI-S STAI-T (all self-report)	T0 = 1 day prior to treatment T1 = 4 weeks post treatment T2 = 6 months post treatment	negative correlations of peak experience and outcome measures	4 dropouts reasons: (2) died of their cancer (2) too ill to continue participating no serious adverse effects
Bogenschutz et al. (2015) psilocybin AUD *N* = 10 **open label**	1st moderate dose (0.3mg/kg) 2nd high dose (0.4mg/kg) optional	2 dosing sessions 4 weeks apart & 12 sessions of psychosocial intervention	all outcomes in tables of study	PACS AASE (all self-report)	T0 = within 2 weeks prior treatment T1 = 1 week post 2nd treatment T2 = 4 months post 2nd treatment	negative correlations of peak experience and outcome measures	1 dropout reason: data not usable no serious adverse effects

Griffiths et al. (2016) psilocybin anxiety & depression in cancer patients *N* = 51 RCT (cross-over)	high dose (0.3-0.4 mg/kg) low dose (0.01-0.04 mg/kg) as placebo	2 dosing sessions 5 weeks apart & PT with crossover: treatment group becomes control group	all outcomes in tables of study	BDI STAI-S STAI-T (all self-report)	T0 = 1 month prior treatment T1 = 5 weeks post 2nd treatment T2 = 6 months post 2nd treatment	negative correlations of peak experience and BDI & STAI-T, but not STAI-S	10 dropouts reasons: (5) disease-progression (1) anxiety (1) vomiting (1) family reasons (1) not returned call no serious adverse effects
Ross et al. (2016) psilocybin anxiety & depression in cancer patients *N* = 29 RCT (cross-over)	moderate dose (0.3 mg/kg) placebo: niacin	1 dosing session 7 weeks apart & PT with crossover: treatment group becomes control group	no outcomes in tables of study email sent to Ross (02.09.22) data won with **WebPlot Digitizer**	BDI STAI-S STAI-T (all self-report)	T0 = 2-4 weeks prior treatment T1 = 6 weeks post 2nd treatment T2 = 6,5 months post 2nd treatment	negative correlations of peak experience and BDI & STAI-Trait, but not STAI-State	6 dropouts reasons: (4) disease - progression (1) died of their cancer (1) resumption of medication no serious adverse effects

Carhart-Harris et al. (2017) psilocybin resistant MDD *N* = 20 **open label**	high dose (0.36 mg/kg) low dose (0.14 mg/kg) as safety dose	2 dosing sessions 7 days apart. no PT, but psychological support	all outcomes in tables of study extension from pilot study by Carhart-Harris et al. (2016) with FU data from pilot study included	BDI STAI-T (all self-report)	T0 = day of treatment T1 = 1 week post 2nd treatment T2 = 6 months post 2nd treatment	negative correlations of peak experience and BDI	1 dropout reason: experience was "blissful but overwhelming" no serious adverse effects
Davis et al. (2021) psilocybin resistant MDD *N* = 27 RCT	1st moderate dose (0.28 mg/kg) 2nd high dose (0.43 mg/kg) delayed treatment as control	2 dosing sessions 1 week apart & 16 PT-sessions compared to delayed treatment condition group 8 weeks afterwards	all outcomes in supplementary material no data for delayed treat group after intervention	BDI-II STAI-S STAI-T (all self-report)	T0 = 3 weeks prior treatment T1 = 4 weeks post 2nd treatment **no T2 data** email sent to Davis (16.08.22)	negative correlations of peak experience and BDI	2 dropouts reasons: (1) sleep difficulties (1) marked reduction in symptoms after 1st session no serious adverse effects

Carhart-Harris et al. (2021) psilocybin resistant MDD *N* = 59 RCT	moderate/high dose (0.36 mg/kg) low dose (0.01mg/kg) as placebo **comparison to escitalopram (SSRI)**	2 dosing sessions 3 weeks apart & 6 weeks of daily SSRI-Placebo vs. 2 placebo sessions 3 weeks apart & 6 weeks of daily escitalopram	all outcomes in tables of study	QIDS-SR-16 BDI – 1A (all self-report)	T0 = 1 day prior treatment T1 = 3 weeks post 2nd treatment **no T2 data** email sent to Carhart-Harris (24.08.22)	negative correlations of peak experience and QIDS-SR-16 results in study of Murphy et al. (2022)	1 exclusion from analysis reason: discontinued taking placebo without informing team no serious adverse effects
Mithoefer et al. (2010) MDMA resistant PTSD *N* = 20 RCT	1st high dose (1.8 mg/kg) 2nd optional supplemental dose (0.9/kg) placebo: lactose	2 dosing sessions 3-5 weeks apart& 11 PT-sessions placebo-group could get treatment after blind was lifted (*N* = 7)	all outcomes in tables of study	CAPS (clinician-administered) IES-R (self-report)	T0 = 4 weeks prior treatment T1 = 3-5 days post 2nd treatment T2 = 8 weeks post 2nd treatment	no measure	2 dropouts reasons: (1) relapse of depression (1) unwillingness to travel no serious adverse effects

Oehen et al. (2013) MDMA resistant PTSD *N* = 12 RCT	high dose (1.8 mg/kg) & optional supplemental dose (0.9/kg) low dose (0.36 mg/kg) as placebo	3 dosing sessions 1 week apart & 12 PT-sessions placebo-group could get treatment after blind was lifted (*N* = 4)	all outcomes in tables of study	CAPS (clinician-administered) PDS (self-report)	T0 = 4 weeks prior treatment T1 = 3 weeks post 3rd session **no T2 data** email sent to Oehen (08.09.22)	no measure	2 dropouts reasons: (1) adverse effects after 1st session (1) died of cancer no serious adverse effects
Danforth et al. (2018) MDMA social anxiety in autistic adults *N* = 12 RCT	1st subgroup moderate dose (1.43 mg/kg) & optional supplemental dose (0.9/kg) 2nd subgroup high dose (1.8 mg/kg) & optional supplemental dose (0.9/kg) placebo: lactose	2 dosing sessions 4 weeks apart & 9 PT-sessions placebo-group could get treatment after blind was lifted (*N* = 4)	all outcomes in tables of study	BDI-II STAI-S STAI-T (all self-report)	T0 = day of treatment T1 = 4 weeks post 2nd treatment T2 = 6 months post 2nd treatment	no measure	1 dropout reason: enrolled with hypertension & pre-study substance use disorder no serious adverse effects

Mithoefer et al. (2018) MDMA resistant PTSD *N* = 26 RCT	high dose (1.8 mg/kg) moderate dose (1.07 mg/kg) low dose (0.43 mg/kg) as placebo	2 dosing sessions 3-5 weeks apart & 12 PT-sessions placebo-group could get treatment after blind was lifted (*N* = 7)	all outcomes in tables of study	CAPS-IV (clinician-administered) BDI-II (self-report)	T0 = day of treatment T1 = 4 weeks post 2nd treatment T2 = 12 months post 2nd treatment	no measure	1 dropout in low dose group no reason reported 1 serious adverse event possibly related to treatment: acute increase in ventricular contractions (recovered fully)
Ot'alora et al. (2018) MDMA resistant PTSD *N* = 28 RCT	high dose (1.8 mg/kg) moderate dose (1.43 mg/kg) low dose (0.57 mg/kg) as placebo	2 dosing sessions 4 weeks apart & 3 PT-sessions & daily telephone contact placebo-group could get treatment after blind was lifted (*N* = 5)	all outcomes in tables of study	CAPS-IV (clinician-administered) BDI-II (self-report)	T0 = day of treatment T1 = 4 weeks post 2nd treatment **no T2 data** email sent to Ot'alora (08.09.22)	no measure	3 dropouts no reasons reported no serious adverse effects

Monson et al. (2020) MDMA resistant PTSD *N* = 12 **open label** MDMA-facilitated **CBCT** (6 couples)	1st moderate dose (1.07 mg/kg) 2nd high dose (1.43 mg/kg) & optional supplemental doses	2 dosing sessions 2-3 weeks apart & 15 CBCT-sessions (condensed version of the 15-session CBCT protocol)	all outcomes in tables of study	CAPS-5 (clinician-administered) BDI-II (self-report)	T0 = 1 week prior treatment T1 = 4 weeks post 2nd treatment T2 = 6 months post 2nd treatment	no measure	no dropouts no serious adverse effects
Wolfson et al. (2020) MDMA anxiety in patients with LTI *N* = 18 RCT	high dose (1.8 mg/kg) & optional supplemental dose (0.9/kg) placebo: lactose	2 dosing sessions 2-4 weeks apart & 9 PT-sessions placebo-group could get treatment after blind was lifted (*N* = 5)	all outcomes in tables of study	STAI-T (main outcome) STAI-S BDI-II (all self-report)	T0 = day of treatment T1 = 4 weeks post 2nd treatment T2 = 6 months post 2nd treatment	no measure	1 dropout reason: disease progression no serious adverse effects

Wang et al. (2021) MDMA severe PTSD *N* = 37 **open label** (multi-site)	**USA** 1st moderate dose (1.14 mg/kg) 2nd & 3rd high doses (1.8 mg/kg) **Canada** 1st moderate dose (1.43 mg/kg) 2nd & 3rd high doses (1.8 mg/kg)	3 dosing sessions 3-5 weeks apart & 12 PT-sessions main objective: train new co-therapist dyads for larger 'phase 3' trials	all outcomes in tables of study .	CAPS-5 (clinician-administered)	T0 = day of treatment T1 = 8 weeks post 3rd treatment **no T2 data** email sent to Wang (09.09.22)	no measure	1 dropout reason: mild increase of nightmares 1 serious event: aborted suicide attempt (probably unrelated to MDMA because history of suicide attempts)
Mitchell et al. (2021) MDMA severe PTSD ***N* = 90** RCT **'phase 3' trial** (multi-site)	moderate dose (1.14 mg/kg) & optional supplemental dose (0.6/kg) high dose (1.8 mg/kg) & optional supplemental dose (0.9/kg) inert placebo	3 dosing sessions 4 weeks apart & 12 PT-sessions	no outcomes in tables of study email sent to Mitchell & official request to MAPS (16.08.22 & 13.10.22) data won with **WebPlot Digitizer**	CAPS-5 (clinician-administered) BDI-II (self-report) (no BDI-Score for T1)	T0 = 1 week prior treatment T1 = 3 weeks post 2nd treatment T2 = 2 months post 3rd treatment	no measure	4 dropouts reasons: (2) COVID (1) distress from adverse effect (1) feeling cured no serious adverse effect

Note: N = sample size; RCT = randomized controlled trial; PT = psychotherapy; MDMA = 3,4-Methylenedioxymethamphetamine; AUD = alcohol use disorder; MDD = major depressive disorder; PTSD = post-traumatic stress disorder; T0 = baseline timepoint; T1 = first assessment-timepoint; T2 = follow-up assessment-timepoint; FU = follow-up; BDI = Beck Depression Inventory; STAI-S = State-Trait Anxiety Inventory-State; STAI-T = State-Trait Anxiety Inventory -Trait; PACS = Penn Alcohol Craving Score; AASE = Alcohol Abstinence Self-Efficacy Scale; QIDS SR-16 = Quick Inventory of Depressive Symptomatology; CAPS = Clinician-Administered PTSD Scale; IES-R = Impact of Event Scale; PDS = Posttraumatic Diagnostic Scale; CBCT = cognitive-behavioral conjoint therapy.