Forside

Eksamensinformation
S650000102 - S650000102: K08: Speciale, Silas Aagaard Andersen

Besvarelsen afleveres af
Silas Aagaard Andersen
siand21@student.sdu.dk

Eksamensadministratorer
Anne-Christina Nielsen
acnielsen@health.sdu.dk
+4565502782

Bedømmere
Maria Semkovska
Eksaminator
msemkovska@health.sdu.dk
+4565504299

Kristine Rømer Thomsen
Censor
krt.crf@psy.au.dk
+4529916092

Besvarelsesinformationer

Titel, engelsk: Effectiveness of MDMA-assisted Psychotherapy for Common Mental Disorders: A Systematic Review
Antal tegn: 104843
Tro og love-erklæring: Ja
Indeholder besvarelsen fortroligt materiale: Nej
Må besvarelsen gøres til genstand for udlån: Ja
Må besvarelsen bruges til undervisning: Ja

Effectiveness of MDMA-assisted Psychotherapy for Common Mental Disorders: A Systematic Review

Master's Thesis
University of Southern Denmark
APA 7 format
May 2023

Author: Silas Aagaard Andersen
Birthdate: July 1st 1991

Instructor: Maria Semkovska

Contents

Resume (dansk)

Introduktion: Forskningsspørgsmålet for dette systematiske review var: *hvor effektiv er MDMA-assisteret Psykoterapi (MP) til at reducere symptomer i voksne patienter med almene psykiatriske diagnoser, og hvor holdbar er effekterne*. Der redegøres for relevansen af dette spørgsmål ved en gennemgang af prævalens- og behandlingsresistens-data for patienter med almene psykiatriske diagnoser, og ved en gennemgang af effektiviteten af traditionel psykoterapeutisk og farmakologisk behandling. Der redegøres ligeledes for rationalet om, at MP sandsynligvis er en ekstraordinær effektiv behandling via MP's psykoterapeutiske og neuropsykologiske teorier.

Metode: Der blev søgt i fire online databaser for kliniske open-label og randomiserede, kontrollerede studier, der anvendte MP som primær intervention til behandling af voksne med almene psykiatriske diagnoser. Effektstørrelser, i form af procentvis symptom reduktion og Cohen's d, blev ekstraheret eller beregnet på baggrund af studiernes tilgængelige data. Derudover blev der ekstraheret demografiske-, studiekarakteristika-, blindings-, komorbiditet-, og behandlingsresistens-data for at vurdere effektstørrelsernes validitet og generaliseringsgrad.

Resultater: Der blev fundet 11 originale kliniske MP studier, der inkluderede forsøgspersoner med diagnoserne PTSD, social angst, og angst, med et total antal forsøgspersoner (*N* = 264). Der blevet fundet store effektstørrelser i de fleste af studierne, men deres validitet blev også fundet usikre. De fleste af de double-blindede studier havde problemer med at opretholde blindingen af forsøgspersonerne og terapeuterne.

Diskussion: Signifikansen af resultaterne diskuteres med særligt fokus på problemerne med opretholdensen af blindingen med inddragelse af placeboteori. Begrænsningerne af de inkluderede studier og af review-processen diskuteres også. Til sidst redegøres der for resultaternes implikationer, der primært drejer sig om behovet for innovation i forskningsdesign, til at løse blindingsproblemerne i MP-studierne, og til at øge forskningshastigheden og udvidelsen af den nuværende, marginale evidensbase.

Abstract

Introduction: This systematic review aimed to answer: *how effective and durable is MDMA-assisted Psychotherapy (MP) in reducing symptoms in adult patients with common mental disorders*. The relevance of this aim was established with reference to prevalence data on common mental disorders, treatment-resistance, and the effectiveness of traditional psychotherapy and pharmacotherapy. The rationale for MP as a potential effective new treatment, was primarily argued for in light of MP's psychotherapeutic and neuropsychological theoretical aspects.

Methods: Four online databases were systematically searched for open-label and randomized controlled studies that had MP as its experimental intervention for any common mental disorders. The primary outcome data were average symptom reductions, and Cohen's d effect sizes, for both within-group and between-group comparisons. Additionally, demographical-, study characteristics-, blinding success-, comorbidity-, and treatment-resistance-data were extracted from each of the included studies, to assess the validity and generalizability of the effects.

Results: 11 original clinical MP studies were found, that included participants with PTSD, social anxiety, and anxiety, with a total sample size of *N* = 264. Large effect sizes were observed for most of the included studies. Low blinding success was observed in most of the double-blinded studies, raising questions about the validity of the effect size estimates.

Discussion: The significance of the results is discussed with particular focus on the low blinding success observed, informed by placebo- and expectancy-theory. Limitations of the included studies and of the review process is also discussed. Finally, the results' implications are discussed, primarily revolving around how future innovation in research designs might help alleviate the problems with blinding observed thus far, and how they might accelerate the rate of research for the currently small evidence base.

Keywords: methylenedioxymethamphetamine, MDMA-assisted psychotherapy, treatment-resistance, effectiveness, efficacy, durability

Abbreviations: 3,4-methylenedioxymethamphetamine (**MDMA**), MDMA-assisted psychotherapy (**MP**), post-traumatic stress disorder (**PTSD**), gross domestic product (**GDP**), Multidisciplinary Association for Psychedelic Studies (**MAPS**), eye movement desensitization and reprocessing (**EMDR**), odds ratio (**OR**), attention deficit hyperactivity disorder (**ADHD**), cognitive behavioral therapy (**CBT**), treatment as usual (**TAU**), serotonin (**5-HT**), selective serotonin reuptake inhibitor (**SSRI**), long-term follow-up (**LTFU**), randomized controlled trial (**RCT**)

Effectiveness of MDMA-assisted Psychotherapy for Common Mental Disorders: A Systematic Review

Introduction Overview

MDMA-assisted Psychotherapy (MP) is an experimental hybrid intervention, mixing pharmacotherapy and psychotherapy in a unique way. In current MP treatment protocols, 3,4-Methylenedioxymethamphetamine (MDMA) is ingested by the patient during 2-3 psychotherapy sessions, each session one month apart. Facilitated by two therapists, the sessions last for about 8 hours. Theory suggests that MDMA's acute effects enhances therapeutic receptivity in patients. Thus far, researchers have mostly focused on MP as a potential treatment for post-traumatic stress disorder (PTSD). The aim of this systematic review was to determine how effective MP is in treating various common mental disorders including PTSD. The reason for this expanded review of disorders is twofold. Firstly, MP-trials for PTSD have already observed great efficacy in treating otherwise treatment-resistant patients. While PTSD is a prevalent disorder, there are several other more prevalent disorders i.e., anxiety, depression, insomnia, and substance abuse. Treatment-resistance is common for all these disorders which potentiates great humanitarian and economic costs, and consequently there is an urgent need for more effective treatments. Secondly, the MP treatment protocol and theory, as well as the psychological effects of MDMA, implies that MP might be effective for non-PTSD disorders as well. As such, it is the aim of this systematic review to evaluate the effectiveness of MP for all these disorders. The following remainder of the introduction elaborates upon the prevalence of common mental health disorders, their current treatments, why MP potentially is a versatile and effective treatment, and some methodological context that is particularly important for this review.

1.1: The Problem: A Mental Health Crisis

1.1.a: Humanitarian Burden

Before exploring MP, it is first investigated what problem MP is supposed to alleviate. Naturally, the overarching problem that attracts all psychotherapy research and development is mental disorders. These disorders are prevalent, often stigmatizing to those who are diagnosed (Kaushik et al., 2016), and are a leading cause for disability (Trautmann et al., 2016). In their meta-analysis, with persons (N = 829,673) drawn from 63 countries across the globe, Steel et al. (2014) found that 17.6%, 95% CI [16.3, 18.9] met diagnostic criteria for a common mental disorder at some point during a 12-month period. They also found that the corresponding life-time prevalence was 29.2%, 95% CI [25.9, 32.6]. Given that the 12-month prevalence is large and that it is more than half of the life-time prevalence, the prevalence statistics suggest a reality in which a significant amount of people live many years of their lives with the burden of mental

suffering. Specifically for Europe, Wittchen et al. (2011) estimated that the 12-month prevalence for common mental diagnoses was 27.1%. That is an estimated 117 million people who are directly affected by mental disorders every year in Europe. The most common mental health disorders, by 12-month prevalence, were anxiety (14.0%), depression (6.9%) drug abuse (>4%), Insomnia (3.5%), and PTSD (2.3%). These statistics suggest that the total humanitarian burden of mental suffering is significant. To have a large impact, interventions should be effective in treating these most prevalent disorder categories.

1.1.b: Economic Burden

The humanitarian burden should be reason enough, in itself, to incentivize the development of solutions to the mental health crisis. However, there is also a significant economic burden that further calls for solutions. In a comprehensive health care report, OECD (2018) estimate that Europe's yearly costs related to mental disorders amount to 600 billion euros, or 4% of GDP. The estimated costs are aggregate costs of treatment, social security-related expenses, and lost labor. Denmark was the European country with the highest costs, in terms of GDP, at 5.2%. Thus, the costs are of significant size and yet they are probably underestimated to some degree. Indeed, the costs of secondary illness caused by stress, depression, drug abuse etc., are not accounted for in these estimates, hence the true costs are probably higher. Furthermore, due to mental disorders' debilitating nature and that people are generally projected to live longer in the future, the costs associated with mental disorders are expected to rise (Trautmann et al., 2016). Suffice to say, the mental health crisis is currently both a heavy humanitarian- and economic burden to society.

1.1.c: Treatment-resistance

To fully understand the societal mental health crisis, one should not only grasp the magnitude of the problem, but also the cause of it. Solutions would surely be easier to find if we understood why there are such high incidence and prevalence rates for mental disorders. However, to exhaustively attempt analyzing and explaining the causes behind the mental health crisis in society, one would have to extensively synthesize evidence from many different fields of expertise, like biology, neuroscience, psychology, psychiatry, philosophy, anthropology, pedagogy etc., (Li et al., 2021; Wolde, 2022) which is of course beyond the scope of this review. The present systematic review considers only part of the problem, which is that currently available treatments are not effective enough to treat all patients. In other words, part of the problem is that a significant minority of mental health patients are treatment-resistant to current treatments. Evidence supporting the existence of treatment-resistance now follows.

Definitions of treatment-resistance varies both within and between disorder types. Several researchers have reviewed the definitions and epidemiology of treatment-resistance for depression. While there was no consensus in the literature as to what constitutes a case of treatment-resistant depression, it

was most often defined as non-responding patients who have been treated with at least two different pharmacological agents (Carvalho & McIntyre, 2015; Voineskos et al., 2020). To classify what constitutes a treatment-response, many have used a 50% symptom reduction cutoff score, to dichotomously divide cases in terms of treatment-response vs. treatment-resistance (Berlim & Turecki, 2007; Nuñez et al., 2022; O'Reardon & Amsterdam, 2001). Several treatment-resistant-depression reviews estimated the proportion of patients who are treatment-resistant to be 10%-40%, all based on the same two primary studies. The first primary study, a prospective follow up study (*N* = 431), found that 12% of the patients, all diagnosed with major depressive disorder, did not remit, as they were treated with antidepressants and psychotherapy for up to 5 years (Keller et al., 1992). The second primary study, a multi-stage intervention trial for outpatients with nonpsychotic major depressive disorder (N = 3,671), found that 33% of the patients did not remit, having been treated with up to four different antidepressants in sequence (Rush et al., 2006). In a separate PTSD-treatment-resistance study with British combat veterans (N = 960) being treated with a mixture of group- and individual cognitive behavioral therapy, it was found that 27% were treatment-resistant with little to no reduction in symptoms (Murphy & Smith, 2018).

The above-described primary studies reported quite different proportions of treatment-resistance. This difference was probably due to heterogeneity in working definitions, patient groups, intervention specifics, follow-up periods etc. These studies, and many others, did not explicitly state all such methodological parameters that may be causes for heterogeneity. It is also important to note that dichotomizing response cutoff scores at 50%, or at an arbitrary remission threshold, leaves out important information about the degree of treatment-resistance. For example, a 45% reduction in symptoms might be considered a partial clinical success, while it would still be coded as a non-response in treatment-resistance studies using 50% cutoff scores. Moreover, a 45% reduction would be counted the same as a 5% reduction. Thus, the convention of dichotomously measuring treatment-resistance restricts more precise estimation. Nevertheless, the aforementioned studies on treatment-resistance and the high prevalence of mental disorders, points to a reality in which there is a lack of adequate, available treatments. Now, serving as a point of comparison for MP, the currently available, conventional treatments and their effectiveness are described.

1.2: Conventional Treatments

1.2.a: Pharmacotherapy

In order to put MP into context, it is necessary to gain a brief overview of the status-quo treatments, which are pharmacotherapy and psychotherapy. The first method, pharmacotherapy, has exclusively in its toolbox an array of psychiatric drugs like antidepressants, antipsychotics, benzodiazepines, stimulants, and more. Descriptive statistics of Denmark's entire patient population is readily available, and

so will serve as a sample country for the prevalence of use. In Denmark, between 2017-2021, the yearly average number of persons prescribed psychiatric drugs, was 733,470, or 12.5% of the population (DHDA, 2021). Out of persons with psychiatric drug prescriptions, 57.8% were prescribed antidepressants and 35.5% were prescribed benzodiazepines, while less were prescribed antipsychotics and stimulants. The number of people who were prescribed multiple drugs was not available. In any case, it is safe to say that antidepressants and benzodiazepines were the most widely prescribed classes of psychiatric drugs. Four large meta-analyses (N = 116,477; 47,950; 13,338; 9,510, respectively) found the efficacy of antidepressants and benzodiazepines for the treatment of depression, insomnia, and anxiety, in terms of odds ratios (OR) vs. placebo, to be in the range (*OR* = 1.5 – 2.2) (Cipriani et al., 2018; De Crescenzo et al., 2022; Kong et al., 2020; Zhou et al., 2020). Although Cohen's d does not directly convey information about absolute differences, Cohen's d is at least an equally important effect size measure to ORs, since Cohen's d weighs cases of 5% symptom reduction and 45% symptom reduction differently, which ORs does not. Using the raw data from the (Cipriani et al., 2018) meta-analysis, Moncrieff (2018) found that the difference in outcomes was equal to *d* = 0.30, which is a relatively small effect size. In addition to small effect sizes, downsides to the use of psychiatric drugs are a range of side effects (Oliva et al., 2021), prescription regimes that require daily ingestion, tolerance buildup, and that patients may develop long-term drug dependency (Gøtzsche, 2016; Soyka, 2017).

The theories as to why psychiatric drugs should work, are generally theories that set neurochemical imbalance as the cause for various mental suffering (Whitaker, 2007). As an example, the serotonin hypothesis describes depression in terms of an epigenetically caused serotonin deficit (Albert et al., 2012), stating that the primary cause for depression is low levels of serotonin. Most antidepressants e.g., selective serotonin reuptake inhibitors (SSRIs), increase the synaptic availability of serotonin, and thus, in theory, alleviates depression. While antidepressants clearly have an effect on the serotonin system (Carhart-Harris & Nutt, 2017) and may alleviate depression to some degree through secondary effects, the serotonin hypothesis is not well supported. In their systematic review of the serotonin hypothesis, Moncrieff et al. (2022) did not find evidence in support of the theory. Patients with depression could not be reliably identified by biomarkers, contrary to what the serotonin hypothesis predicted. Neuroscientists do not yet grasp all the dynamics of the serotonin system, and in turn the psychological effects of antidepressants are not yet fully understood. As is later shown, there are good evidence to suggest that short-term elevation of serotonin levels has a positive effect on mood. However, long-term elevation of neurotransmitter levels can cause tolerance build-up, such as with the frequent long-term use of antidepressants. Increased tolerance may reduce the drug's effectiveness and may lead to dependence (Gøtzsche, 2016). So, even if antidepressants are effective short-term, as usually measured with short follow-up periods, they are

probably not effective as long-term solutions. Imbalance theories are also prevalent beyond those for serotonin, e.g., dopamine hypotheses for schizophrenia and attention deficit hyperactivity disorder (ADHD), but are not be pursued further here. The point made here, in light of chemical imbalance theories in general and the nature of tolerance build-up, is that long-term use of psychiatric drugs is not likely to maintain durable effects for many mental health patients.

1.2.b: Psychotherapy

The second method, by which mental disorders are most often treated, is talk psychotherapy. There are many schools of psychotherapy with their own theories and approaches. However, there is a set of common factors which are part of most psychotherapies (Wampold et al., 1997; Wampold, 2015). The most important of these common factors, in terms of explained variance of effect, is therapeutic alliance. Constituted by three parts, therapeutic alliance is the patient's and therapist's (1) agreement on goals, (2) agreement on methods, and (3) empathic bond or level of trust. Furthermore, specific disorders may have important treatment principles that all of the specific and most effective therapies have in common e.g., trauma-focused exposure therapy for PTSD. Generally, the most researched and applied school of psychotherapy is cognitive-behavioral therapy (CBT) (Dragioti et al., 2017). It is often found that psychotherapy combined with pharmacological treatment is more effective than either alone, and that psychotherapy is slightly more effective than pharmacological treatment, considering adherence to treatment (Cuijpers et al., 2013; Cuijpers et al., 2020). An umbrella-review for 21 disorders (N = 137,126), estimated that the efficacy of pharmacotherapy and psychotherapy was d = 0.50 (Huhn et al., 2014). However, as a critique of the former umbrella review's inclusion of wait-list controls, and lack of statistical weighting, Leichsenring et al. (2022) conducted a more stringent and extensive umbrella-review (N = 650,514), and also included pharmacotherapy comparisons. The latter umbrella-review found the following effect sizes; psychotherapy vs. placebo or treatment as usual (TAU), d = 0.34; pharmacotherapy vs. placebo or TAU, d = 0.36; and psychotherapy vs. pharmacotherapy, d = 0.11. With the perspective of decades of psychotherapy efficacy research, the reviewers concluded that there was a ceiling effect for psychotherapy's effectiveness, and that a paradigm shift in such research seemed necessary. Another independent umbrella-review, including 247 meta-analyses, found that only 16 (7%) of those meta-analyses were without bias and provided convincing evidence for the efficacy of psychotherapy (Dragioti et al., 2017). In sum, based on the presented umbrella-reviews, the efficacy of psychotherapy seems to be in the range of d = 0.3 – 0.5.

To summarize thus far, as the premise for this review, mental health disorders are widely prevalent and are costly in terms of individuals' quality of life and societal resources. Current treatments, that is pharmacotherapy and psychotherapy, seem to work well for some patients, but not for all. Meta-analyses

of the conventional treatments' effectiveness suggest that a ceiling has been reached and that a paradigm shift is needed to develop more effective therapies. MP is innovative not only because it incorporates the use of MDMA, but also in the way that MP incorporates MDMA. Rather than being guided by chemical imbalance theories, MP proponents suggest that MDMA can open a therapeutic 8-hour window of opportunity, to facilitate lasting change that does not require years or a lifetime of therapy. To understand what MP is and why MP is potentially extraordinarily effective for PTSD and non-PTSD disorders alike, it is helpful to understand how MP developed and what MDMA is.

1.3: MDMA: What is it?

1.3.a: Brief Historical Context of MDMA and MP

3,4-Methylenedioxymethamphetamine (MDMA) only exists synthetically and was first formulated by the German pharmaceutical company Merck in 1912. MDMA was not proposed as a potential psychotherapeutic agent before Alexander Shulgin synthesized and ingested it in 1978 (Freudenmann et al., 2006). Shulgin wrote down his impressions from his first personal uses of MDMA, illustrating some of the subjective effects (Shulgin & Shulgin, 1991):

"(with 120 mg) I feel absolutely clean inside, and there is nothing but pure euphoria. I have never felt so great, or believed this to be possible. The cleanliness, clarity, and marvelous feeling of solid inner strength continued throughout the rest of the day, and evening, and through the next day. I am overcome by the profundity of the experience, and how much more powerful it was than previous experiences, for no apparent reason, other than a continually improving state of being. All the next day I felt like 'a citizen of the universe' rather than a citizen of the planet, completely disconnecting time and flowing easily from one activity to the next."

As Shulgin and others shared such personal accounts in the late 70's, MDMA spread as a psychotherapeutic and recreational drug. Due to its recreational use and perceived dangers, it was made illegal worldwide in 1984-1985. In response to the prohibition, Rick Doblin, who already had experience with and belief in MDMA's psychotherapeutic potential, founded the Multidisciplinary Association for Psychedelics Studies (MAPS) in 1986. Based in USA, MAPS then formulated MP for PTSD treatment protocols. Rick Doblin was thus primarily responsible for developing MP. Important to note, with regards to MP potentially being effective for non-PTSD diagnoses, is that Rick Doblin has openly stated in interviews (Elton, 2019) that his choice of focusing on PTSD were as much a political reason as it was a scientific one. His political argument was that in working with PTSD, he would be able to work with veterans whom are particularly well respected in American society. Thus, working with veterans served as a political counterbalance to the stigma of funding research on an illicit psychedelic drug. Then, in 2017, after two

decades of safety and pilot studies, MP was given Breakthrough Therapy Designation (FDA, 2017) due to promising signs of efficacy, sparking significant additional research interest.

1.3.b: MDMA's Neuropharmacology

Ingested orally, MDMA acutely causes very high levels of serotonin (5-HT), and to a lesser extent, high levels of dopamine, norepinephrine, and oxytocin (Amoroso, 2015). For the purpose of understanding why MDMA may enhance psychotherapy, it is helpful to briefly review how MDMA interacts with each affected neurotransmitter system. It is also illustrative to compare how antidepressants relate to the serotonin system specifically. All antidepressants and MDMA have the common characteristic that they increase 5-HT availability, albeit MDMA does so at much higher levels and for a shorter time. The half-life of the most popular antidepressants, SSRIs, typically is about 1 day (Andrade, 2022), whereas MDMA's half-life is about 8 hours for the 100 mg clinical doses (Papaseit et al., 2016). MDMA and various types of antidepressants causes higher 5-HT synaptic availability in unique ways. SSRIs blocks the 5-HT reuptake transporter on the pre-synaptic neuron, causing higher 5-HT availability in the synaptic cleft. MDMA, on the other hand, enters the presynaptic neuron through the reuptake transporter and similarly blocks the reuptake of 5-HT, but furthermore reverses the transporter function to effectively flush out stored 5-HT from the pre-synaptic neuron into the synaptic cleft (Green et al., 2003). In this way, MDMA facilitates very high levels of 5-HT availability. Although MDMA is similar enough in structure to 5-HT itself, that it can enter the pre-synaptic neuron through the reuptake transporters, MDMA is not compatible with any of the post-synaptic receptors, like 5-HT is. In other words, MDMA does not mimic neurotransmitters, but rather facilitates the release of endogenous neurotransmitters. Since the presence of 5-HT in the synaptic cleft is much higher with the use of MDMA, the acute 5-HT-related psychological effects are also much more pronounced, compared to antidepressants. MDMA also directly affects dopamine, norepinephrine, and oxytocin levels. While some antidepressants also affect norepinephrine or dopamine in conjunction with serotonin, no antidepressants do so to the extent that MDMA does, and they never affect oxytocin. This is to say that MDMA has a very unique neuropharmacological profile. MDMA causes higher levels of dopamine and norepinephrine, principally the same way it causes higher levels of 5-HT i.e., by entering the presynaptic neurons, blocking reuptake, and flushing the endogenous neurotransmitters into the synaptic cleft (Green et al., 2003). However similar in mechanism, the extent of release is greater for 5-HT than for dopamine or norepinephrine. Lastly, MDMA effectively causes higher levels of the hormone oxytocin, yet it is not well understood by what mechanism (Vizeli & Liechti, 2018).

1.3.c: MDMA's Psychological Effects

Many of the psychological effects of MDMA, and of antidepressants, are primarily mediated by the serotonin system. The psychological functions of the serotonin system as a whole is far from fully

understood, but empirical evidence carries therapy-relevant implications (Carhart-Harris & Nutt, 2017). Long-term tolerance aside, acutely increasing 5-HT availability tends to cause a decrease in anxiety, fear, stress, impulsivity, and aggressive behavior. Dopamine and norepinephrine are both catecholamines, are similar in structure, and serve synergistic psychological functions, facilitating wakefulness, focus, euphoria, cognition, motivation, and more. Some of the psychological effects of higher levels of oxytocin are increased openness for social interaction and attachment. In sum, MDMA promotes (1) decreased anxiety, fear, stress, impulsivity, and aggressive behavior, (2) increased cognition, motivation, wakefulness, focus, and euphoria, and (3) increased openness to attachment and social interaction. The suggested psychological effects of MDMA are derived from animal studies, but are nevertheless consistent with effects self-reported by humans (Baylen & Rosenberg, 2006), although the effects may vary between individuals, between uses, and within uses, depending on contextual factors. Important for clinical implications, and central to the popular argument that MDMA is useful for PTSD interventions, are the effects of decreased aggression, anxiety, and fear. Patients diagnosed with PTSD tend to experience intense anger, anxiety, and fear, that may be a hindrance for treatment (Hinton et al., 2022). However, decreased anxiety and fear, or increased wakefulness and feelings of euphoria for that matter, could also easily be argued to be beneficial for patients diagnosed with depression disorders. Many evidence-based psychotherapies operate under principles of gradual exposure, cognitive restructuring, and attachment styles. All these principal treatment processes are arguably enhanced by the effects of MDMA, suggesting a broader therapeutic application of MDMA than for just PTSD. By virtue of MDMA's seemingly versatile and therapy-conducive psychological effects, MP may be generally effective in the treatment of many disorders, hence the purpose of this review. Treatment guidelines and supporting theories from MP are now described, to further argue that MP is potentially effective in treating mental health patients with other diagnoses than PTSD.

1.4: MDMA-Assisted Psychotherapy: How Does it Work?

1.4.a: Practical Aspects

To gain some basic insight on the practical aspects of MP, a clinical MP-trial for the treatment of PTSD have been reviewed (Mitchell et al., 2021). The treatment program spanned 18-20 weeks, having three experimental MDMA-augmented psychotherapy sessions, each with a duration of 8 hours. Before, between, and after each of the drug-augmented sessions, there were three 90-minute talk therapy sessions to help prepare and debrief the drug-augmented sessions. Thus, the treatment program had a total of about 40 hours of therapy. All sessions included two therapists, one from each sex. The participants were lying comfortably in a bed and were afforded eyeshades and headphones with music, and was encouraged to attend to the sensory experience. Participants were also encouraged to intermittently take off their

headphones and share their experience with the therapists. Part of the therapist-role-doctrine was to be non-directive and let the participants lead the topics of conversation. Thus, participants alternated between focusing on their inner experience and sharing and discussing the experience with the therapists. In support of the premise of the present review, it should be noted that none of these practical aspects excludes MP's applicability for most non-PTSD disorders. The theoretical aspects of MP are now reviewed in the same light.

1.4.b: Fundamental Theoretical Aspects of MP

As outlined in the MP treatment protocol (MAPS, 2010), a fundamental theoretical assumption of MP, is that the patient's mind and brain, in a trusting, safe, and therapeutic environment, have an innate ability to make salient to the patient's consciousness what is necessary for healing. As has been argued, MDMA helps facilitate such an environment. Essentially, however, the protocol also states as its basic premise, that the therapeutic effect is heavily dependent on conventional psychotherapeutic common factors e.g., therapeutic alliance. It is further stated that this fundamentally important synergistic effect proposedly emerges through the triple interaction of (1) MDMA's neuropsychological effects, (2) the therapeutic setting, and (3) the mindsets of patients and therapists. Firstly, it has already been argued in the present review that the neuropsychological effects of MDMA is likely beneficial to psychotherapy more generally than for just PTSD. Secondly, the therapeutic setting, interchangeably known as therapeutic alliance, is widely regarded as the most important common factor of all psychotherapies (Barkham et al., 2021). The therapeutic alliance has various definitions, most of which emphasize trust and bonding between therapist and patient. For the current argument, the most important thing to note, is that therapeutic alliance is a common factor, important for virtually all psychotherapies across diagnoses. Thirdly, consideration of set and setting is also not specific to PTSD-treatment, but is rather a general preparatory checklist when using psychedelic substances such as MDMA. *Set* refers to the internal state of the patient, such as mindset, personality, mood, knowledge, convictions, and intentions. *Setting* refers to the physical, social, and cultural environment in which the therapy takes place. Thus, the stated three most important treatment principles in MP are not specifically tailored to PTSD, but rather seems to be part of a non-specific psychotherapy framework. This supports the premise that MP may be effective for non-PTSD disorders as well.

1.4.c: Important Elements of MP

The treatment manual further specifies 14 therapy-method elements that are important to MP as a treatment for PTSD (MAPS, 2010). Only two elements are specifically addressing trauma, which is at the heart of the PTSD diagnosis. The first of those two elements, is the principle of gradual trauma-focused exposure. This treatment principle is central to all conventional evidence-based PTSD-therapies

(Mavranezouli et al., 2020). The second element is that therapist support, such as nurturing touch, guided breathing etc., should be available in case of somatic manifestations of trauma. This second element functions to prevent overexposure, supporting the former element. These two elements can potentially be substituted with other disorder-specific treatment principles, when MP is applied to disorders like depression, substance abuse, and so on. Nothing in the remaining 12 essential therapy-method elements were specific to PTSD, and moreover, most of them were conservative compared to conventional psychotherapeutic theory. Categorically, these remaining elements concerned (1) safety, (2) therapist experience, (3) knowledge of MDMA effects, (4) therapeutic alliance, (5) preparation and integration, (6) non-directive therapy, (7) evoking content, and (8) set and setting.

In sum, MP seems to be a predominantly generic and non-specific psychotherapy framework. There were very little in the practical and theoretical aspects of MP to suggest incompatibility as a treatment for anxiety, depression, insomnia, substance use etc. Arguably, MP may be versatile in its application, in the same way cognitive behavioral therapy is widely applied, with relatively little modification between a wide range of disorders. Thus, the objective of this review was primarily to find out how effective MP is for common mental disorders. To preface the methods section of this systematic review, several important methodological factors in reviewing MP is first contextualized.

1.5: Important Factors to Consider When Evaluating MP's Effectiveness

To see what important factors to include in the evaluation of MP's effectiveness, a hypothetical, ideal study is first described and commented upon. So, to gain maximum confidence in MP's effectiveness, studies would ideally show and include (1) large standardized mean differences e.g., Cohen's d, (2) large relative reductions of symptoms, (3) no treatment drop-out, (4) perfectly maintained placebo blinding, (5) participants with previously treatment-resistant, long-lasting mental health diagnoses, (6) participants with no prior use of or expectancies about MDMA, (7) long-term follow-up (LTFU), and (8) large sample sizes. A standardized mean difference, such as Cohen's d, can be deceptive on its own, as small reductions in symptoms with low variance, can show large effect sizes. Therefore, to buttress Cohen's d, a symptom reduction measure that better represents the clinical significance of the effect should also be shown. Drop-out rates are important to include because they imply treatment failure from a conservative standpoint, and are usually omitted from the studies' effect sizes calculations. For reference, in their meta-analysis Swift and Greenberg (2014) estimated that 19.2% on average drops out of various psychotherapy treatments for depression, based on studies k = 161. Similarly for PTSD, with k = 92, it was estimated that 21.0% dropped out. The LTFU are important to assess durability of the treatment, which is an essential part of the rationale for MP i.e., intense short-term therapy with long-term durable effects.

With regards to placebo, some important terms are defined here. A *placebo* is a sham treatment, usually a pill or treatment procedure, that is designed to convince control group patients that they received the true treatment (Kirsch, 2009). *Placebo effects* are therapeutic responses in participants (Mitsikostas & Benedetti, 2019), caused by (1) participant expectations of the true treatment and (2) participant degree of belief in actually receiving the true treatment, and (3) differential treatment by therapists due to bias. The true treatment, also called active treatment, is in this context MP with full-dose MDMA. The placebo-control treatment, is either MP with inert placebo pills or MP with low-dose MDMA. Finally, the degree of expectation of improvement, is defined as *expectancy* and *expectancy level*. The purpose of a placebo control group is to allow differentiation of the true treatment effects from the placebo effects. This differentiation is entirely dependent on participants and therapists not knowing the group to which the participants have been allocated. Thus, it is important to include measures of blinding success and expectancy levels, especially when the treatment involved has salient effects, as is the case for MP. As a reference for blinding success, Lin et al. (2022) in their meta-analysis with $k = 16$, found that about 40% of participants in antidepressant studies incorrectly guess their treatment allocation, where 50% is the ideal.

There are two important factors to consider, such that study outcomes are not entirely invalidated as a consequence of low blinding success. Firstly, although it is not common practice to measure expectancy levels in participants, it may be important information if blinding success is low. If the expectancy levels are known, the placebo effects can potentially be estimated and subtracted. This of course is not a perfect solution, since therapist bias is not accounted for, but it enables a partial correction to the inflation of effect sizes that the placebo-effects may produce. Secondly, one of the premises of this review, is that a significant number of patients are treatment-resistant to current treatments, thus making the search for new treatments relevant. To assess whether MP is effective in treating treatment-resistant patient populations, it is important to include demographic measures of treatment resistance. Treatment-resistance measures is also important, when comparing effect sizes to other clinical studies, that may have included patient populations with different levels of treatment-resistance. As such, blinding success, expectancy levels, and treatment-resistance, are important factors to consider when qualifying the effectiveness of MP.

1.6: Rationale and Aims for this Systematic Review

Current unmet needs in mental health care makes the search for effective treatments important. MP is potentially an effective treatment for otherwise treatment-resistant patients. To date, all systematic MP-reviews on effectiveness have only included PTSD as the target disorder in their reviews (Bahji et al., 2019; Illingworth et al., 2021; Smith et al., 2022; Tedesco et al., 2021). It has been argued that this singular association with PTSD is mostly theoretically and politically based. In reviewing the basic principles of MP

and the effects of MDMA, it seemed that MP also has great potential in treating other disorders than PTSD, and so MP was systematically reviewed as such. The previous systematic MP-reviews were further limited, in that they did not systematically report quantified outcomes of blinding success, nor did they report participant expectancy levels. The previous systematic reviews reported LTFU outcomes, but several of the reviewed studies were missing LTFU data at the time. To address these limitations of the previous systematic reviews, PTSD was again included in the scope of the present review, in addition to all common mental disorders. Open-label studies and randomized controlled trials (RCTs) with low blinding success are rightfully considered inferior to RCTs with high blinding success. However, as described, studies with low blinding success can still be meaningful sources of evidence, given adequate consideration of factors such as expectancy levels and treatment-resistance.

Consequently, the present review aimed to answer the primary research question (1) *how effective and durable is MP in reducing symptoms in adult patients with common mental disorders?* This research question was sought answered by systematically extracting and analyzing data from original open-label-, RCT-, and LTFU-MP-studies. The primary outcome variables for this systematic review were symptom reduction scores, relative symptom reduction, and Cohen's d effect sizes. Firstly, within-group outcome change scores were differentiated by the initial treatment-exit scores and the LTFU scores. Secondly, for studies with a control group, between-group effect sizes were calculated, one at treatment-exit, and another at LTFU. Thus, a potential four sets of effect sizes were calculated for each study.

Additionally, to scrutinize the primary outcome variables of effectiveness adequately, the aim was to operationalize and extract relevant contextual variables that could answer the related questions (1a) *what were the general study characteristics?* (1b) *what were the general demographics of the participants?* (1c) *how successful was blinding*? (1d) *what level of expectancy did the participants have*? and (1f) *what was the participants' previous level of treatment-resistance?* The variables answering these questions, and how they were operationalized, are defined in the methods section.

Methods

Prisma guidelines for conducting and reporting systematic reviews were followed (Moher et al., 2009).

2.1: Study Inclusion and Exclusion Criteria

Only peer-reviewed studies that were published in English were included. The patient population were any adult (18+) patients of any sex that met the diagnostic criteria for one or more of the following mental health disorders: major depressive disorder (MDD), any anxiety disorder, bipolar disorder, schizophrenia, obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), any personality

disorder, attention deficit hyperactivity disorder (ADHD), any substance use disorder, any panic disorder, any phobia disorder, anorexia nervosa or any other eating disorder, or suicidality. Patients were diagnosed by the guidelines of DSM-III, DSM-III-R, DSM-IV, DSM-IV-TR, DSM-5, ICD-9, ICD-10 or ICD-11. To be included, the intervention studies had to use MDMA-assisted Psychotherapy as their active treatment arm and had to quantify pre- and post-treatment outcomes with a relevant symptom scale e.g., the Clinician-Administered PTSD Scale (CAPS) as a measure of PTSD. Studies that reported otherwise-relevant outcome measures, but only as secondary outcomes measures, were excluded. Both open-label and RCTs were eligible. Studies were allowed to include any number of hours of preparatory-, integratory-, and active MP-treatment sessions, above a minimum of one active MP-treatment session, in which the participant were dosed with MDMA, at any dosage above 50mg MDMA. Studies were allowed to include any number of participants. The RCTs control group design was allowed to be either traditional inert placebo controls or active controls i.e., with a dose below 50mg of MDMA. Follow-up studies to the primary studies were also included.

2.2: Database Search

The electronic databases PsycArticles, PsycINFO, PubMed, and Embase were searched from their inception to February 2023. For each database the following search string was used: (open label OR pilot OR phase 1 OR phase 2 OR phase 3 OR randomized OR rct OR efficacy OR effectiveness OR intervention OR therapy OR treatment OR trial).ti.ab. AND (disorder OR depression OR anxiety OR PTSD OR schizophrenia OR anorexia nervosa OR phobia OR suicid* OR abuse OR diagnos*) AND (MDMA OR 3,4-methylenedioxy-methamphetamine OR 3,4-methylenedioxymethamphetamine).

2.3: Study Selection- and Data Collection -Process

The data collection and study selection were performed by one rater. All searched database studies were imported to reference managing software Endnote Version 20 and duplicates were removed. Firstly, titles and abstracts were priority screened against the inclusion/exclusion criteria, for potential full text review. Secondly, studies were selected by comparing their described methods against the inclusion/exclusion criteria. Thirdly, the data was extracted directly and manually from the selected studies.

2.4: Extracted Variables

Here the extracted variables are categorized under either the primary research question, or one of the related questions, according to their primary relevance.

(1) ***How effective and durable is MP in reducing symptoms in adult patients with common mental disorders?*** (a) Sample size i.e., an aggregate of the active treatment group and the control group. The sample sizes were exclusively a count of participants who at least received one active- or one controlled- 8-hour MP-session. (b) Time-point of outcome measures i.e., how many months had past at the time of

measurement, since the final MP-treatment session. (c) Two pairs of drop-out rates, no. and percentages, differentiated by participants dropping out between the baseline- and treatment-exit-time-points and participants dropping out between the treatment-exit- and LTFU-time-points. To qualify as a drop-out, the participant had to receive at least one MDMA-dosed MP-session and then discontinue. Drop-outs due to death with no reasonable causal link to the studies, were not counted as drop-outs. (d) Baseline-, treatment-exit-, and LTFU-scores of the primary outcome variables, mean (SD), were differentiated by active treatment group and control group. (f) Based on the extracted outcome scores, Cohen's *d* [95% CI] and average symptom reduction percentages, was calculated as shown at 2.5: Statistical and Analytical Methods.

(1a) ***What were the general study characteristics?*** (a) Author and year. (b) Study design, with possible coding; open-label/cross-over/randomized/placebo-controlled/active-controlled/triple-blind/double-blind. Open-label were coded for studies that never included a control-group. Cross-over were coded for studies in which the control group at some point crossed to full-dose active treatment. Controlled studies were categorized by either having an inert placebo-control group or a low-dose MDMA-control group, both of which otherwise included therapists and psychotherapy protocols equal to the active treatment groups. Double-blind was coded for controlled studies in which both participants and therapists did not know to which group the participant had been allocated. If additionally, the outcome raters were also blinded, triple-blind was coded. (c) Country, where the study was conducted. (d) MAPS-sponsorship (yes/no). (f) Primary target diagnosis of the study e.g., PSTD. (g) Primary outcome scale e.g., CAPS for PTSD. (h) MDMA-assisted psychotherapy sessions, no. (i) MDMA dosage in the full-dose active treatment group, mg. (j) An optional 50% extra dose was offered during the MP-sessions (yes/no). (k) MDMA dosage in the low-dose MDMA-control group, mg.

(1b) ***What were the general demographics of the participants?*** (a) Females in the sample, no. and percentages. (b) Age, years, mean (SD). (c) Participant epidemiology. Any relevant epidemiological participant commonalities were noted e.g., victims of sexual assault/abuse. (d) Past or present comorbidities, no. and percentages, limited to the diagnoses; anxiety, depression, insomnia, substance abuse including alcoholism, and suicidal behavior.

(1c) ***How well was blinding maintained?*** (a) Blinding, therapists/raters/participants i.e., who were blinded to group allocation. (b) Blinding evaluation, therapists/raters/participants, i.e., who were asked to guess participants' group allocation. (c) Incorrect guesses, no. and percentages. The incorrect guesses were differentiated by four possible case combinations, i.e., person guessing (therapist/participant) and person's group allocation (active/control).

(1d) ***What level of expectancy did the participants have*****?** (a) The study quantified expectancy levels from participants (yes/no). (b) Standardized expectancy measure, e.g., credibility/expectancy questionnaire (CEQ). (c) Participant prior use of MDMA (yes), no. and percentages. (d) Prior use of MDMA in the last five years (yes), no. and percentages.

(1f) ***What was the participants' previous level of treatment-resistance?*** (a) Diagnosis lifetime in years, mean (SD). (b) Participants that previously received therapy, no. and percentages, by four categories; psychotherapy, psychiatric medications, anxiolytics, and antidepressants.

2.5: Statistical and Analytical Methods

When available, the sample standard deviations were calculated from the raw data, and otherwise the studies' reported standard deviations were used. For each study, Cohen's d effect sizes and confidence intervals were calculated with the weighted pooled standard deviation of the compared scores. The within-group average symptom reduction effect sizes were calculated as a ratio of the change scores and the baseline scores.

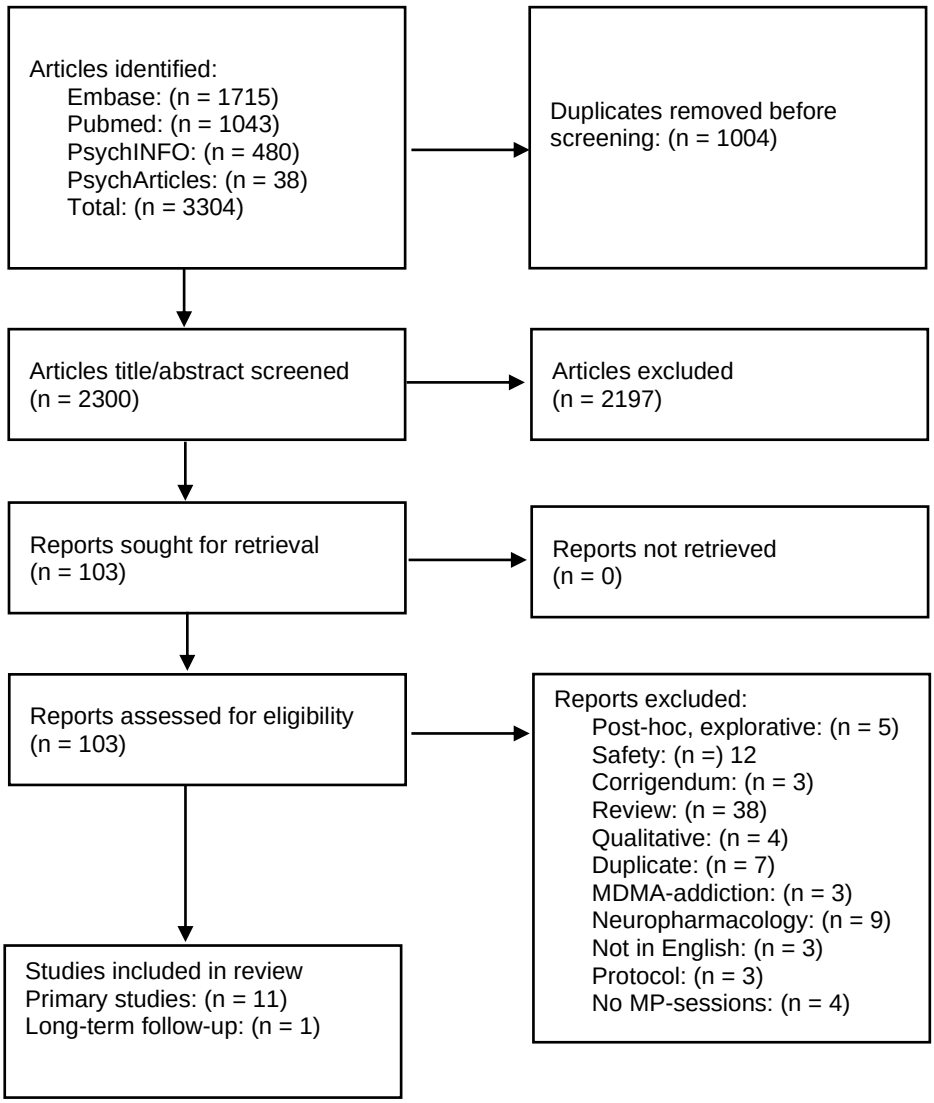
A combined quantitative and qualitative approach were used to analyze and scrutinize the quality and magnitude of the evidence. All the contextual variables were analyzed to qualify the validity of the primary outcome scores. For Cohen's d between-group effect sizes, the conventional standard was used for comparison (Leppink et al., 2016), that is d = 0.2, d = 0.5, and d = 0.8, being small-, medium-, and large-effect sizes, respectively. Similarly, each treatment groups' average relative symptom reduction was compared against the conventional clinical response cutoff scores at 30-50% (Nøhr et al., 2021; Papakostas et al., 2020) meaning that symptom reduction group averages above 30-50% were considered clinically significant.

Results

The search strategy identified 2300 unique articles, 103 of which were isolated as potentially relevant for this review, as shown in Figure 1. After assessment, 12 studies were included in the review, one of which (Mithoefer et al., 2013) reported LTFU outcome data to one of the 11 original studies (Mithoefer et al., 2011). In addition to the one LTFU study, six of the 11 original studies directly provided LTFU within-group outcome data. Only one study reported LTFU between-group outcome data. The total sample size was (*N = 264)*.

Figure 1

Database Search and Study Selection



3.1: Study Characteristics

The individual studies' characteristics have been summarized in Table 1 and Table 2. All 12 studies had been published between 1991 and 2021, predominantly in the USA. All studies were sponsored and guided by MAPS, pointing to homogeneity in the treatment approaches, but also to potential bias. For their target diagnoses, nine studies targeted PTSD, one targeted social anxiety in people with autism (Danforth et al., 2018), and one targeted anxiety in people with life threatening illnesses (Wolfson et al., 2020). All PTSD-studies used CAPS-4 or CAPS-5 as their primary outcome measure, except one using the Severity of Symptoms Scale for Post-Traumatic Stress Disorder (SSSPTSD). The social anxiety study used the Liebowitz

Social Anxiety Scale (LSAS), and the anxiety study used the trait-part of the State-Trait Anxiety Inventory scale (STAI). Three of the 11 studies were open-label, whereas the remaining eight were all randomized, controlled, and with raters, therapists, and participants being blinded to group allocation, also known as triple-blind. Three of the controlled studies used low-dose MDMA-pills for their control group, ranging from 25-40mg dosing, while the remaining five studies used inert placebo pills for their control groups. Five out of eight of the controlled studies, let the control-group-allocated participants cross over to active treatment after the first outcome measure time-point, with the implication that LTFU between-group comparisons were not possible for those studies. The LTFU time points, as seen in Table 7, Table 8, and Table 9, ranged 6-45 months after the last MP-treatment. For all studies, the first primary outcome scores were measured 0-2 months after treatment-exit. The treatment in all studies consisted of 2-3 MP-sessions with 75-125mg doses with accompanying preparatory and integratory sessions, except for (Bouso et al., 2008) which only offered one MP-session dosed at 50-75mg. All but two studies offered participants a 50% extra booster dose 90-120 minutes after the initial dose (Bouso et al., 2008; Danforth et al., 2018). Mitchell et al. (2021) had a sample size, *n* = 90, while the remaining studies had relatively small sample sizes between (*n* = 3) and (*n* = 37).

Table 1

Study Characteristics

Author (year), sample size	Study design	Country	MAPS Sponsorship	Target Diagnosis	Primary Outcome Scale
Jardim et al. (1999), *n* = 3	OL	Brazil	yes	PTSD	CAPS-4
Bouso et al. (2008), *n* = 6	TB, R, PC,	Spain	yes	PTSD	SSSPTSD
Mithoefer et al. (2011), *n* = 20	TB, R, PC, C	USA	yes	PTSD	CAPS-4
Oehen et al. (2013), *n* = 12	TB, R, APC, C	Switzerland	yes	PTSD	CAPS-4
Danforth et al. (2018), *n* = 12	TB, R, PC	USA	yes	Social Anxiety	LSAS
Mithoefer et al. (2018), *n* = 26	TB, R, APC, C	USA	yes	PTSD	CAPS-4
Ot'alora et al. (2018), *n* = 28	TB, R, APC, C	USA	yes	PTSD	CAPS-4
Monson et al. (2020), *n* = 12	OL	USA	yes	PTSD	CAPS-5
Wolfson et al. (2020), *n* = 18	TB, R, PC, C	USA	yes	Anxiety	STAI (Trait)
Mitchell et al. (2021), *n* = 90	TB, R, PC	USA	yes	PTSD	CAPS-5
Wang et al. (2021), *n* = 37	OL	USA/Canada	yes	PTSD	CAPS-5

Note. OL = Open-label; TB = Triple blind; R = Randomized; PC = Placebo-controlled; APC = Low-dose MDMA active placebo-controlled; C = Crossover placebo group to open-label active treatment, after the first primary outcome post-measure; MAPS = Multidisciplinary Association for Psychedelic Studies; CAPS = Clinician Administered PTSD Scale; SSSPTSD = Severity of Symptoms Scale for Post-Traumatic Stress Disorder; LSAS = Liebowitz Social Anxiety Scale (LSAS); STAI = State-Trait Anxiety Inventory (STAI).

Table 2

Study Characteristics – Continued

Author (year), sample size	MDMA-assisted sessions, no.	Dosage MDMA, mg	50% extra dose	Active placebo dosage, mg
Jardim et al. (1999), *n* = 3	3	75-125	yes	n/a
Bouso et al. (2008), *n* = 6	1	50-75	no	no
Mithoefer et al. (2011), *n* = 20	2-3	125	yes	no
Oehen et al. (2013), *n* = 12	3	125	yes	25
Danforth et al. (2018), *n* = 12	2	75-125	no	no
Mithoefer et al. (2018), *n* = 26	2	75-125	yes	30
Ot'alora et al. (2018), *n* = 28	2-3	100-125	yes	40
Monson et al. (2020), *n* = 12	2	75-100	yes	n/a
Wolfson et al. (2020), *n* = 18	2	125	yes	no
Mitchell et al. (2021), *n* = 90	3	80-120	yes	no
Wang et al. (2021), *n* = 37	3	80-100	yes	n/a

Note. 50% extra dose refers to participants being offered a booster dose of MDMA 90-120 minutes after the initial dose.

3.2: Demographics and Comorbidities

The demographics have been summarized in Table 3. On average, females made up 62.1% of the participants, an expected majority for many studies of PTSD, when caused by sexual assault/abuse. In two studies, one with veterans of war and first responders (Mithoefer et al., 2018) and one with people with autism (Danforth et al., 2018), the majority of participants were men, which is concordant with those participant populations at large. The mean age of each of the participant samples ranged from 31 to 55 years of age with standard deviations between *SD* = 7 to *SD* = 13. For the seven studies that reported the participants' prior MDMA/Ecstasy use, it was found between 0%-55.6% had previously used MDMA in their lifetime, and a maximum of 21.1% had used it within the last five years before the start of the study. The PTSD studies' participants were victims of sexual assault/abuse, veterans of war, first responders, and

victims of childhood neglect, where most studies had a mix of such participants. The study targeting social anxiety included only people with autism (Danforth et al., 2018), whereas the study targeting anxiety included only people with life-threatening illnesses (Wolfson et al., 2020). Uniquely, Monson et al. (2020) included six couples, being six men and six women, whom were treated for PTSD in a couples-therapy-adjusted MP modality.

Table 3

Participant demographics

Author (year), sample size	Females, no. (%)	Age, years, mean (SD)	Prior MDMA/Ecstasy use		Participant characteristics
			yes, no. (%)	Within last 5 years, no. (%)	
Jardim et al. (1999), *n* = 3	2 (66.7)	40.3 (5.0)	n/a	0 (0.0)	Sexual Assault/Abuse
Bouso et al. (2008), *n* = 6	6 (100)	39 (n/a) *	n/a	n/a	Sexual Assault/Abuse
Mithoefer et al. (2011), *n* = 20	17 (85.0)	40.4 (7.2)	9 (45.1)	n/a	Sexual Assault/Abuse
Oehen et al. (2013), *n* = 12	10 (83.3)	41.4 (11.2)	1 (8.3)	n/a	Various
Danforth et al. (2018), *n* = 12	2 (16.7)	31.3 (8.8)	0 (0.0)	n/a	Autism
Mithoefer et al. (2018), *n* = 26	7 (26.9)	37.2 (10.3)	6 (23.1)	n/a	Veterans/First responders
Ot'alora et al. (2018), *n* = 28	19 (67.9)	42.0 (12.9)	n/a	n/a	Various
Monson et al. (2020), *n* = 12	6 (50.0)	47.1 (12.5)	n/a	n/a	Various/Couples therapy
Wolfson et al. (2020), *n* = 18	14 (77.8)	54.9 (7.9)	10 (55.6)	2 (20.0)	LTI
Mitchell et al. (2021), *n* = 90	59 (65.6)	41.0 (11.9)	29 (32.2)	≤19 (21.1)	Various
Wang et al. (2021), *n* = 37	22 (59.5)	35.6 (10.8)	n/a	n/a	Various

Note. * = age estimated by reported age ranges. LTI = Life threatening illness.

The participant history of comorbidities has been summarized in Table 4. Five studies reported proportions of participant historical comorbidity, ranging between 7.7%-83.3%. The majority of studies (9 out of 11) reported high comorbidity with depression, ranging between 66.7%-100%, showing that depression was the most prevalent disorder in the participants' mental health histories, besides their primary disorder. For the four reporting studies, anxiety was highly prevalent in two of those studies, at 75.0%-83.3%. Only one study reported insomnia comorbidity (Wolfson et al., 2020), at 61.1%. Six studies reported on Substance Abuse, including alcohol abuse. The reports varied between 0%-33% of participants

having had some type of substance abuse. Finally, five studies included prevalences on suicidal behavior, with reported prevalences between 16.7%-32.2%. The suicidal behavior was coded via the Colombia-Suicide Severity Rating Scale (C-SSRS). Although several studies did not report on some or all of the possible comorbidities, their data suggests that participants most often had a present or past comorbid mental disorder.

Table 4

Participant History of Comorbidities

Author (year), sample size	Participant Comorbidities, no. (%)				
	Anxiety	Depression	Insomnia	Substance Abuse*	Suicidal Behavior
Jardim et al. (1999), *n* = 3	n/a	3 (100)	n/a	n/a	n/a
Bouso et al. (2008), *n* = 6	n/a	n/a	n/a	n/a	n/a
Mithoefer et al. (2011), *n* = 20	3 (15.0)	16 (80.0)	n/a	2 (10.0)	n/a
Oehen et al. (2013), *n* = 12	n/a	10 (83.3)	n/a	3 (25.0)	n/a
Danforth et al. (2018), *n* = 12	2[1] (16.6)	8 (66.7)	n/a	≥2 (16.7)	2 (16.7)
Mithoefer et al. (2018), *n* = 26	2 (7.7)	20 (76.9)	n/a	n/a	11 (42.3)
Ot'alora et al. (2018), *n* = 28	n/a	21 (75.0)	n/a	5 (17.9)	8 (28.6)
Monson et al. (2020), *n* = 12	9 (75.0)	10 (83.3)	n/a	4 (33.3)	n/a
Wolfson et al. (2020), *n* = 18	15 (83.3)	14 (77.8)	11 (61.1)	n/a	3 (16.7)
Mitchell et al. (2021), *n* = 90	n/a	82 (91.1)	n/a	n/a	29 (32.2)
Wang et al. (2021), *n* = 37	n/a	n/a	n/a	0 (0.0)	15 (40.5)

Note. Suicidal behavior was categorized with scores ranging from lifetime scores 1-5 on the Colombia-Suicide Severity Rating Scale (C-SSRS). * = alcohol abuse included. 1 = Generalized anxiety disorder (the primary diagnosis was social anxiety).

3.3: Previous Treatment Resistance

Extracted variables relating to treatment resistance have been summarized in Table 5. Five studies reported the average duration of the participants' diagnoses to be 5.7 to 29.4 years on average. With the exception of (Wang et al., 2021) and (Oehen et al., 2013), all studies reported that 75%-100% of their participants had received previous psychotherapy. Most participants, although with some uncertainty in (Mitchell et al., 2021), reported having used previous psychiatric medication. Antidepressants were the most prevalent at 58.3% to 96.2%, while anxiolytics (hereunder benzodiazepines) were reported to have

been used in 25.0%-88.5% of cases. The reported diagnosis lifetimes and the previously received therapy-rates points to a very high degree of treatment resistance in the included studies' participant populations.

Table 5

Participant History of Treatment Resistance

Author (year), sample size	Diagnosis lifetime, mean (SD), years	Participants that previously received therapy, no. (%)			
		Psychotherapy	Psych. Meds.	Anxiolytics.	Antidepressants
Jardim et al. (1999), *n* = 3	n/a	3 (100) *	3 (100) *	n/a	n/a
Bouso et al. (2008), *n* = 6	n/a	6 (100) *	6 (100) *	n/a	n/a
Mithoefer et al. (2011), *n* = 20	20.7 (14.4)	15 (75.0)	15 (75.0)	n/a	n/a
Oehen et al. (2013), *n* = 12	18.3 (12)	n/a	n/a	n/a	n/a
Danforth et al. (2018), *n* = 12	n/a	≥10 (83.3)	≥7 (58.3)	3 (25.0)	7 (58.3)
Mithoefer et al. (2018), *n* = 26	5.7 (1.25)	25 (96.2%)	25 (96.2)	23 (88.5)	25 (96.2)
Ot'alora et al. (2018), *n* = 28	29.4 (19.3)	28 (100)	n/a	15 (53.6)	20 (71.4)
Monson et al. (2020), *n* = 12	n/a	11 (91.7)	n/a	n/a	n/a
Wolfson et al. (2020), *n* = 18	n/a	18 (100)	n/a	n/a	n/a
Mitchell et al. (2021), *n* = 90	14.0 (11.0)	88 (97.8)	≥17 (18.9)	n/a	n/a
Wang et al. (2021), *n* = 37	10.6 (11.5)	1 (2.7)	n/a	19 (51.4)	27 (73.0)

Note. * = participants had used either psychotherapy or pharmacotherapy, but it was not reported whether they had used both.

3.4: Blinding Success

The blinding success variables have been summarized in Table 6. All of the eight controlled studies were triple-blinded, but only six of them asked both the participants and the therapists to guess group allocations. For successful blinding, approximately 40%-60% incorrect guesses are expected, with 50% being ideal. For all studies, therapists assigned to the active treatment group guessed incorrectly only about 0%-14.0% of guesses, suggesting that blinding was generally not successful. Incorrect guesses were more frequently made by the participants than the therapists, with 41.9% in the (Ot'alora et al., 2018) study and 33.3% in the (Oehen et al., 2013) study, both studies utilizing low-dose MDMA for their control groups. The third and final study that utilized low-dose MDMA for their control group was (Mithoefer et al., 2018), whom reported that therapists guessed incorrectly about 40.7% of the time and participants 53.7%. However, these latter percentages in (Mithoefer et al., 2018) were not included in Table 6, as they were

conflated with incorrect guesses between two active treatment conditions (75mg vs. 125mg) and the 30mg control group, making blinding success unclear for that study.

Compared to the studies with inert-placebo-pill control groups, the studies with low-dose MDMA controls had superior blinding success. Generally, compared to the active treatment groups, there was slightly higher blinding success in the control groups, based on both therapist and participants guesses. However, as the number of guesses were generally low in each condition, there were a high degree of uncertainty as to the precision of the results. It is observed that (Ot'alora et al., 2018) had achieved the most successful blinding condition, possibly due to utilizing a 40mg MDMA dose for their control group. The worst blinding, with only 1 incorrect guess out of 20 possible, was reported by (Mithoefer et al., 2011) whom utilized an inert placebo pill for their control group. Generally taken, there were relatively low blinding success in all but one study (Ot'alora et al., 2018).

Table 6

Blinding Success

Author (year), sample size	Blinding (T/R/P)	Blinding evaluation (T/R/P)	Incorrect guesses, no. (%)			
			Therapists, active c.	Participants, active c.	Therapists, placebo c.	Participants, placebo c.
Bouso et al. (2008), *n* = 6	T, R, P	P, T	1 (12.5)	1 (25.0)	4 (100)	1 (50.0)
Mithoefer et al. (2011), *n* = 20	T, R, P	P, T	0 (0)	1 (5.0) *	0 (0.0)	1 (5.0)
Oehen et al. (2013), *n* = 12	T, R, P	P, T	0 (0)	2 (33.3)	1 (33.3)	2 (50.0)
Danforth et al. (2018), *n* = 12	T, R, P	P, T	4 (12.9)	0 (0)	3 (18.8)	1 (12.5)
Mithoefer et al. (2018), *n* = 26	T, R, P	none	n/a	n/a	n/a	n/a
Ot'alora et al. (2018), *n* = 28	T, R, P	P, T	n/a (14.0)	n/a (41.9)	n/a (22.7)	n/a (27.3)
Wolfson et al. (2020), *n* = 18	T, R, P	P, T	4 (11.1)	n/a	5 (13.9)	n/a
Mitchell et al. (2021), *n* = 90	T, R, P	T ^	n/a	7 (15.9)	n/a	2 (4.3)

Note. T = Therapists. R = Raters. P = Participants. Blinding refers to which persons were blinded, whereas blinding evaluation refers to which persons were asked to guess the given participants' group allocation. * = guesses in both treatment conditions were aggregated, so it is unknown to which group the participant with the correct guess were allocated. ^ = Blinding evaluation information was only gathered anecdotally in (Mitchell et al., 2021).

3.5: Effectiveness at Treatment-exit

The *within-group* effect sizes of MP at 0-2 months post treatment-exit, have been summarized in Table 7. Between the first treatment session and the treatment-exit time point, the weighted average drop-

out rate of all studies was 6.8%, with six of the studies having no drop-out, and the highest drop-out rate being 14.2%. The average drop-out rate was thus very low compared to the usual average around 20%, suggesting very good tolerability and adherence to the treatment. All but two studies showed large and statistically significant effect sizes above Cohen's *d* = 2, having the lower boundaries of their 95% CIs above *d* = 1.10. For these studies, compared to baseline measures, symptoms were reduced on average by 49.5%-68.1%. The two studies with lower effect sizes (Bouso et al., 2008; Oehen et al., 2013) still had effects in the positive direction, but were not statistically significant, including 0 in their confidence intervals. These two studies had average symptom reductions of 23.5%-28.1%. The sample-size-weighted average symptom reduction, between all studies, were 53.8%. The largest within-group Cohen's d effect size of all studies, was found in one of the PTSD studies (Wang et al., 2021) with *d* = 3.25. Interestingly, the two next largest within-group effect sizes was observed in the social anxiety study (Danforth et al., 2018) with *d* = 2.92, and the anxiety study (Wolfson et al., 2020) with *d* = 2.59, suggesting that MP may be effective for non-PTSD anxiety disorders as well.

Table 7

Treatment-exit Within-group Effectiveness

Author (year), sample size ¤	Time-point, months ^	Drop-out, no. (%) *	Baseline Score, Mean (SD)	Outcome Score, Mean (SD)	Avg. Symptom Reduction, %	Effect Size, Cohen's d [CI 95%]
Jardim et al. (1999), *n* = 3	2	0 (0.0)	80.0 (26.9)	32.0 (26.9)	60.0	2.39 [0.30; 4.49]
Bouso et al. (2008), *n* = 4	0	0 (0.0)	40.0 (7.0)	28.6 (8.7)	28.1	1.43 [-0.13; 2.98]
Mithoefer et al. (2011), *n* = 12	2	2 (9.1)	79.4 (23.3)	25.3 (21.5)	68.1	2.41 [1.36; 3.46]
Oehen et al. (2013), *n* = 8	1	2 (14.2)	66.4 (13.6)	50.8 (19.7)	23.5	0.92 [-0.11; 1.95]
Danforth et al. (2018), *n* = 7	1	0 (0.0)	91.8 (15.8)	46.4 (15.2)	49.5	2.92 [1.47; 4.38]
Mithoefer et al. (2018), *n* = 19	1	2 (7.7)	87.0 (17.3)	37.5 (29.1)	56.9	2.07 [1.28; 2.86]
Ot'alora et al. (2018), *n* = 21	0	0 (0.0)	92 (18)	44.6 (28.6)	51.6	2.04 [1.35; 2.74]
Monson et al. (2020), *n* = 12	0	0 (0.0)	41.4 (5.8)	19.4 (13.7)	53.2	2.10 [1.10; 3.09]
Wolfson et al. (2020), *n* = 13	1	0 (0.0)	62.5 (7.3)	38.9 (10.6)	37.8	2.59 [1.55; 3.64]
Mitchell et al. (2021), *n* = 42	2	11 (12.2)	44 (6.0)	19.6 (15.6)	55.5	2.10 [1.58; 2.62]
Wang et al. (2021), *n* = 36	1	1 (2.7)	45.4 (7.2)	15.6 (10.9)	65.7	3.25 [2.55; 3.95]

Note. ^ = Measure time-point refers to how many months since the last active MDMA-assisted therapy session the outcome was measured, rounded to the nearest integer. ¤ = sample sizes shown in this table only include those

participants who had received the active treatment, either originally or via cross-over, and who actually had their outcomes measured. * = drop-out-rates were based on the active treatment group- and the control group-participants that had received at least one treatment.

The *between-group* effectiveness at 0-2 months post treatment-exit have been listed in Table 8. All eight controlled studies showed positive between-group effects. However, four of the studies did not show statistically significant differences, as they included $d = 0$ in their confidence intervals. In terms of raw change scores, all studies showed more than double the average change compared to their control groups, except for (Mitchell et al., 2021) which were just short of double the change score compared to the control group. The lowest between-group effect size was observed in (Ot'alora et al., 2018), even though the change score of the active treatment group was still double that of the control group. The lower effect size for (Ot'alora et al., 2018) compared to the other studies, may partly be a result of high blinding success, as shown in Table 6. As mentioned, blinding was more successfully maintained by the studies utilizing low-dose MDMA for their control groups, compared to the normal placebo-controls. To see if effect sizes were different between studies using low-dose MDMA controls and those using the inert placebo controls, a post-hoc effect size comparison was made. It was found that the studies with low-dose MDMA controls, had a weighted average effect size $d = 1.07$, whereas the studies with inert placebo controls had a correspondent weighted average effect size $d = 1.02$. Thus, the post-hoc analysis suggested, unexpectedly, that the studies with low-dose MDMA controls observed the same level of effectiveness as the studies with inert placebo controls.

Generally, the between-group effect sizes were large, although the confidence intervals were also relatively wide, with all of the lower bounds at $d = 0.79$ or lower. The study with the largest drop-out-subtracted sample size, ($n = 79$) (Mitchell et al., 2021), reported a between-group effect size at $d = 0.91$. Out of all the included studies, the largest between-group effect size was $d = 1.79$ (Mithoefer et al., 2018). In sum, the effect sizes observed in the active treatment groups were large vs. both types of control groups. However, the results were based on relatively small sample sizes with accompanying high degrees of uncertainty.

Table 8

Treatment-exit Between-group Effectiveness

Author (year), sample size ¤	Time-point, months^	Treatment group change score, Mean (SD)	Placebo group change score, Mean (SD)	Effect Size, Cohen's d [CI 95%]
Bouso et al. (2008), *n* = 6	0	11.25 (9.9)	4.5 (2.1)	0.78 [-0.97; 4.02]
Mithoefer et al. (2011), *n* = 20	2	54.1 (34.3)	20.5 (9.4)	1.22 [0.25; 2.20]
Oehen et al. (2013), *n* = 12	1	15.6 (18.1)	3.2 (15.3)	0.72 [-0.52; 1.95]
Danforth et al. (2018), *n* = 11	1	44.1 (15.2)	19.3 (18.8)	1.50 [0.12; 2.88]
Mithoefer et al. (2018), *n* = 26	1	49.5 (23.5)	11.4 (12.7)	1.79 [0.79; 2.78]
Ot'alora et al. (2018), *n* = 27	0	25.5 (27.4)	11.5 (21.2)	0.53 [-0.39; 1.45]
Wolfson et al. (2020), *n* = 18	1	23.5 (13.2)	8.8 (14.7)	1.08 [-0.01; 2.17]
Mitchell et al. (2021), *n* = 79	2	24.4 (11.6)	13.9 (11.5)	0.91 [0.44; 1.37]

Note. ^ = Time-point refers to how many months since the last active MDMA-assisted therapy session the outcome was measured, rounded to the nearest integer. ¤ = sample sizes shown in this table only include those participants who actually had their outcomes measured.

3.6: Durability: Effectiveness at Long-term Follow-up

Seven studies reported LTFU within-group outcome scores, which are listed in Table 9. Two studies followed up at 6 months, four studies at 12 months, and one study at 45 months. The weighted average drop-out rate between the post-measurement time-point and the LTFU time-point, was 3.9%. The symptom reductions observed at LTFU were in the range 45.8%-69.6%. The sample-size-weighted average symptom reduction, between all seven LTFU-studies, were 57.8%. The LTFU symptom reduction scores suggest that the treatment effects were well maintained, and even improved a little, compared to the outcome scores at treatment-exit. All LTFU within-group Cohen's d effect sizes were above *d* = 2, and had all improved since treatment-exit, except for (Danforth et al., 2018) which had declined slightly to *d* = 2.71. Notably, Oehen et al. (2013) saw a large enough symptom reduction since the first treatment-exit measure to become statistically significant at LTFU. The largest increase in within-group effect size, since the initial treatment-exit scores, was found in (Ot'alora et al., 2018), which had increased to *d* = 2.88 up from *d* = 2.04. This latter increase was probably, at least partly, due to the fact that all participants, including those who crossed over, had received three MP-sessions at follow-up, whereas the original active treatment group participants had only received two MP-sessions at the first outcome measure time-point. The increased symptom reductions observed in (Ot'alora et al., 2018) suggest that MP's therapeutic effects cumulate each session to a significant degree, up to at least three MP-sessions. To the primary research

question, the results suggest that the effects of MP are durable at least several months to years after treatment-exit, with slight improvement over time for the majority of the studies.

Table 9

Long-term Follow-up Within-group Effectiveness

Author (year), sample size ¤	Time-point, months ^	Drop-out, no. (%) *	Baseline Score, Mean (SD)	Outcome Score, Mean (SD)	Avg. Symptom Reduction, %	Effect Size, Cohen's d [CI 95%]
Mithoefer et al. (2011), *n* = 16	45	3 (15)	77.9 (20.4)	23.7 (22.7)	69.6	2.51 [1.58; 3.44]
Oehen et al. (2013), *n* = 11	12	0 (0.0)	~	~	41.2	~
Danforth et al. (2018), *n* = 7	6	0 (0.0)	91.8 (15.8)	42.9 (20.4)	53.3	2.71 [1.30; 4.11]
Mithoefer et al. (2018), *n* = 19	12	2 (7.7)	87.1 (16.1)	38.8 (28.1)	55.5	2.11 [1.43; 2.79]
Ot'alora et al. (2018), *n* = 25	12	0 (0.0)	92.0 (18.0)	31.0 (24.2)	66.3	2.88 [2.12; 3.65]
Monson et al. (2020), *n* = 12	6	0 (0.0)	41.4 (5.8)	15.5 (15.2)	62.5	2.25 [1.23; 3.27]
Wolfson et al. (2020), *n* = 17	12	0 (0.0)	61.1 (7.0)	33.1 (11.0)	45.8	3.04 [2.05; 4.02]

Note. ^ = Measure time-point refers to how many months since the last active MDMA-assisted therapy session the outcome was measured, rounded to the nearest integer. In the case of Mithoefer et al. (2011), the time-point was the average months to follow-up, with several months' variance between individuals. * = drop-out rates were based on both the active treatment group- and the placebo control group participants. ¤ = sample sizes shown in this table only include those participants who actually had their outcomes at LTFU. ~ = Oehen et al. (2013) did not report raw scores for LTFU, but reported that the average symptom reduction scores were 28 points and that the difference to baseline was statistically significant.

Only one study (Danforth et al., 2018) reported its LTFU outcomes vs. a control group, since two studies did not report LTFU at all, and six studies had control group-crossover designs. Danforth et al. (2018) reported a Leibowitz Social Anxiety Scale (LSAS) mean (SD) change score of 47.7 (14.7) in the treatment group vs. 23.3 (18.0) in the control group, with effect size *d* = 1.54 [0.15; 2.92]. It was a very large effect size, but must be considered in the light of a wide confidence interval and a low sample size.

Discussion

Treatment-resistance to status-quo treatments is prevalent, so any promising avenues of new psychotherapeutic treatments should be explored and researched. MP seems to be one such treatment, having shown promising results as a treatment for PTSD in treatment-resistant populations, as concluded in previous systematic reviews (Bahji et al., 2019; Illingworth et al., 2021; Smith et al., 2022; Tedesco et al., 2021). However, extraordinary claims of effectiveness require extraordinary evidence and scrutiny thereof. Furthermore, if MP is indeed very effective for PTSD, it may also be very effective for other common diagnoses. Thus, the present systematic review sought to answer the research question: *how effective and durable is MP in reducing symptoms in adult patients with common mental disorders?* To answer this question, outcome data from clinical MP studies were systematically gathered and analyzed. Study design-, demographic-, treatment-resistance-, and blinding success-data were also extracted and analyzed. The summary and meaning of the results are now presented and discussed.

4.1: Results Discussion

The initial, brief answer to the research question, is that MP *may* be an effective and durable treatment for anxiety disorders. Certainly, the symptom reductions and the between-group effect sizes were generally found to be large at treatment-exit and equally so at LTFU. It was no surprise to see the large effects in the PTSD studies, as many of the included studies had already been covered in previous reviews. Interestingly, equally large effects were observed in the study with anxiety due to life threatening illness, and in the study with social anxiety in people with Autism. This suggests that MP may also be effective and durable in treating anxiety disorders more broadly. However, no other disorder type was covered by the included studies. If the symptom reductions and Cohen's d effect sizes were the only considered factors for effectiveness, the conclusion for the included studies would have been that MP is very likely effective. However, effectiveness in its most conservative definition, is the therapeutic effects beyond those of placebo effects, which is now referred to as *true effectiveness* in this review, as opposed to *observed effectiveness* which includes any potential placebo-effects. Thus, while the observed effectiveness was large, there were several findings that created a large degree of uncertainty as to what the true effectiveness of MP might be, in both the positive and negative direction. These factors are now considered in turn, to give proper nuance to the effect size findings.

4.1.a: Factors Threatening Effect Size Validity

(1) The included studies had relatively small sample sizes, with resultant wide confidence intervals. Even though effect sizes were generally large, there was uncertainty of their precision. Nevertheless, from a statistical point of view, the means and averages of the results found are more likely to be the true value of

the effect sizes than any other value. Thus, assuming no other threats to validity, MP is *likely* effective around *d* = 1 vs. control groups, with symptoms reductions around 50-60%. However, there were other threats to validity.

(2) The included studies generally had poor blinding success and unknown expectancy levels. The low blinding success observed turned out to be the greatest threat to the validity of the effect sizes. While the studies that incorporated low-dose MDMA for their control group (Mithoefer et al., 2018; Oehen et al., 2013; Ot'alora et al., 2018) saw some success in blinding therapists and especially participants, it was only partial successes. The studies that utilized inert placebo pills for their control condition (Bouso et al., 2008; Danforth et al., 2018; Mitchell et al., 2021; Mithoefer et al., 2011; Wolfson et al., 2020), generally reported low blinding success. The degree to which this influenced the effect sizes is unknown, but the conservative standpoint is to assume that the effect sizes have been inflated to some degree (Barkham et al., 2021). Furthermore, it is not standardized procedure in psychotherapy research to quantify participant expectancy levels (Muthukumaraswamy et al., 2021), and indeed none of the included studies had explicit measures of expectancy. However, the majority of studies did report prior MDMA-use, which may serve as a rough proxy for expectancy. It may be assumed that people with prior MDMA-use have elevated expectancy levels (Butlen-Ducuing et al., 2023). For those studies that reported prior MDMA-use, it was found that a large minority of the participants had previously tried MDMA. This suggests that at least a significant minority of participants may have had higher than normal expectancy levels. Additionally, even those participants that had not used MDMA previously, may have had high expectancy levels simply due to the media attention and hype around MP or other psychedelic psychotherapies more generally (Bedi et al., 2022). Thus, the lack of direct expectancy measures combined with the low blinding success observed, makes for uncertain conclusions around what proportion of the effect sizes were placebo-driven.

(3) There was a lack of LTFU between-group comparisons, as most studies crossed over their control-group participants to an active group, reportedly due to ethical reasons associated with treating severe and treatment-resistant mental disorders (Mithoefer et al., 2018; Mithoefer et al., 2011; Oehen et al., 2013; Ot'alora et al., 2018; Wolfson et al., 2020). The scientific argument for not crossing over control groups, is to have a measure of how much the control groups would have naturally increased or decreased their symptom levels, methodologically known as maturation and regression to the mean (Barkham et al., 2021). However, the lack of controlled LTFU is overshadowed by the low blinding success observed. In a blinded study with low blinding success in the control group, the value of the control group is severely reduced, which concordantly reduces the value of the LTFU comparison to that group. Furthermore, the participants included in the studies had histories of treatment-resistance, and one could argue that the natural regression of symptoms is lower in treatment-resistant patients. It is tautological that patients with

diagnosis life-times of over 10 years, do not experience significant natural symptom regression. As such, the significance of the lack of LTFU between-group comparisons is somewhat reduced.

(4) A further issue with crossing participants over from control groups to active treatment, were potential testing effects on the LTFU within-group outcome scores. The participants who crossed over to active treatment, received more total hours of therapy than the original active treatment groups. Thus, the symptom reductions in the crossed-over participants may have been greater than for the original active treatment groups. The LTFU scores reported, were aggregates between the original active treatment groups and the subsequent cross-over groups (Mithoefer et al., 2018; Mithoefer et al., 2011; Oehen et al., 2013; Ot'alora et al., 2018; Wolfson et al., 2020). This aggregation may be partly the reason why the within-group effect sizes observed at LTFU were slightly larger than those initially measured at treatment-exit.

4.1.b: Factors Strengthening Effect Size Validity

(1) The participants in the included studies' samples, were a relatively difficult-to-treat group of patients, evident by three observations. Firstly, all studies that reported diagnosis life-times, except one (Mithoefer et al., 2018), reported that the large majority of participants had been living with their diagnoses for more than 10 years. Secondly, all studies reported that the large majority of participants had previously tried psychiatric medication, psychotherapy, or both, except for (Oehen et al., 2013) who did not report this data. Thirdly, only two studies did not report comorbidity demographics (Bouso et al., 2008; Wang et al., 2021). All remaining studies reported that the large majority of participants had a history of comorbidity, with depression being particular widespread. Comorbidity, especially with depression, often decrease the likelihood of successful psychotherapy treatment (Chekroud et al., 2018). Altogether, the three observations suggest that the samples were very difficult-to-treat patient groups. Assuming all else is equal, studies that show effectiveness in difficult-to-treat patient populations, arguably provides stronger evidence than studies without such a population. Thus, the recruitment of difficult-to-treat patient populations in most of the included studies, is considered a favorable factor for the effectiveness of MP.

(2) Drop-out rates were low in all studies, with half of the studies having no drop-out at all (Bouso et al., 2008; Danforth et al., 2018; Jardim et al., 1999; Monson et al., 2020; Ot'alora et al., 2018; Wolfson et al., 2020). As mentioned, it is normal to have drop-out rates around 20% (Swift & Greenberg, 2014), whereas the average drop-out rate for the present review's studies was three times lower, at 6.8%. The uncertainty of an intervention's true treatment effects is proportional with the studies' drop-out rates, with the most plausible cause of drop-out being non-response or intolerable treatment (Bell et al., 2013). Thus, the low drop-out rates observed in the present review's studies, suggest a smaller degree of uncertainty of the effect size estimates, and that it was relatively easy for participants to adhere to the MP treatment protocol.

4.1.c: All Factors Considered

When considering all the above factors, the conclusion is that MP is observed to be effective, but it is nevertheless still unknown if it is actually effective beyond placebo-effects i.e., what the so-called *true effects* are. In spite of the large effect sizes, there is too much uncertainty around the placebo- and expectancy-effects to firmly claim that MP is truly effective. While many of the included participants were difficult-to-treat and with low drop-out rates, these factors do not outweigh the pressing issues of low blinding success and lack of expectancy measures. As pessimistic as that finding appears, the large effect sizes are still encouraging. Exactly because the observed effect sizes are large in a very difficult-to-treatment patient population, it is warranted and needed that researchers seek to solve the issues surrounding placebo and expectancy, to narrow down the uncertainty of the true effectiveness of MP. Furthermore, while it was argued, as part of this review's rationale, that MP might be effective for many different common mental disorders, only two non-PTSD studies were found. As such, there is still an enormous void of evidence that needs to be filled to empirically show the true potential of MP for other disorders. Due to the low blinding success observed, placebo- and expectancy-effects becomes the central issue of this systematic review's discussion. Placebo and other relevant theories are now explored, with the purpose of further contextualizing the results and inspiring future research directions.

4.2: Results in Relation to Placebo- and Other-Existing Theory and Evidence

The available data on MP did not allow for a firm conclusion as to how effective MP is for common mental disorders. Low blinding success is the biggest cause for uncertainty. An argument in favor of MP, in spite of the low blinding success, is that the observed effects are so large that the effects are most likely not only due to placebo effects, given the very difficult-to-treat patient groups. While this is a reasonable argument, it is difficult to support with empirical evidence. It would be natural to compare the MP results to other clinical studies with similarly difficult-to-treat patient populations. However, it is rare for psychotherapy or pharmacotherapy trials to include similarly difficult-to-treat patients in their studies, especially when conducted for commercial purposes like most antidepressant studies are (Cipriani et al., 2018). This commercial bias is unfortunate, since pharmacotherapy trials have much higher blinding success than traditional talk psychotherapy trials (Juul et al., 2021), making pharmacotherapy the best standard comparator with regards to blinding success. Even when the research incentive is purely academic, researchers will often have a bias toward the success of the intervention (Leichsenring et al., 2017). These biases may generally disincentivize the inclusion of participants who have previously shown treatment-resistance with long and comorbid histories of diagnoses. In other words, MP trials have generally treated patients that are not easily comparable with most other psychotherapy research. As such, the primary importance for this systematic review falls on the issue of low blinding success. To understand

why the blinding issue may be critical in evaluating MP's effectiveness, it is helpful to review some first principles of study design- and placebo-theory.

4.2.a: Fundamental Reasons for the Double-blind Randomized Controlled Study Design

The double-blind randomized controlled study design, sometimes referred to as the golden standard, is regarded as one of the most robust designs for inferring causation between an intervention and its outcomes. Randomization refers to randomly allocating participants to active treatment groups or control groups (Barkham et al., 2021). The purpose of randomization is to decrease important variable differences between two compared groups i.e., equalize the expectancy levels, as well as their general demographics, average baseline scores, and what else are considered important differential factors. For any new hyped intervention like MP, there may be selection bias in the included participants (Bedi et al., 2022). Those patients who hear and believe good things about MP, have higher expectancy levels to MP, and are more likely to apply for participation. However, this bias will at least be approximately equal between the two treatment groups within the study, if they have been randomly distributed between the active group and the control group. The purpose of the control group is furthermore to allow differentiation of the true effectiveness from the observed effectiveness (Muthukumaraswamy et al., 2021). The observed effectiveness is comprised by (1) maturation i.e., normal developmental changes that may reduce symptoms, (2) regression to the mean i.e., the tendency for symptoms to regress naturally toward normal symptom levels over time, (3) placebo- and expectancy- effects, (4) investigator/therapist bias interacting with treatment fidelity i.e., therapists treating participants differentially based on group allocation and not due to the planned intervention differences, and of course (5) the true effectiveness of the intervention (Scott et al., 2022). Thus, in theory, the control group allows for differentiation of the true effectiveness from the observed effectiveness. However, placebos and therapist bias are only controlled for and equalized between the two groups, in so far that both participants and therapists are blind to which treatment group they have been allocated i.e., the double-blinded requirement (Kaptchuk, 1998). The present systematic review generally observed low blinding success, which then raises questions as to how much of the observed effects are placebo effects. One may ask, how strong are placebos anyway?

4.2.b: The Strength of Placebos

There is no doubt that placebos are therapeutic, if they are believed by patients to be true treatments (Kirsch, 2009). Placebos have been shown to be mediated by psychological expectations, which in turn affects neurological activity, potentially moderating mood, pain perception, inflammation, and many other psychobiological states (Finniss et al., 2010). A tempting argument to make in favor of MP's effectiveness, is that placebo effects probably can't be durable up to the LTFU time-point. However, there is evidence to suggest that the placebo responses in antidepressant trials are indeed often durable until

LTFU (Khan et al., 2008). Placebos are potentially powerful treatments, but all placebos are not equally powerful. The strength of placebo pills has been shown to depend on several factors such as brand name, color, proclaimed dose, salience of intervention, and more (Kirsch, 2009). More relevant to the present review, placebo strength also depends on how much participants believe in the placebos i.e., expectancies or expectancy levels (Mitsikostas & Benedetti, 2019). This suggests that a pill brand only affects the placebo effects, in so far, the patients have learned to associate the brand with something positive or negative. Negative expectations can give adverse responses, called nocebo, rather than therapeutic placebo responses (Locher et al., 2019). Thus, the strength and effect of a placebo cannot be assumed to be the same between studies, when there are significant differences between the interventions that the placebos are supposed to mimic. In other words, expectancy levels and the consequent placebo effects in an antidepressant trial may be vastly different from those in MP trials, even if blinding is perfectly maintained. One example is that participants are encouraged to lay in a bed during the MP trials, which may induce a sense of receiving more intensive care, as compared to walking home with antidepressant pills in hand to be self-administered. As such, the strength of placebo effects in MP trials cannot be easily inferred from placebo effects in other psychotherapy- or pharmacotherapy-trials.

4.2.c: Nothing but Placebo?

If MP is nothing but a placebo, it can at least be said that MP seems to be a very good placebo, given the large symptom reductions and the difficult-to-treat participants included. Still, one might ask whether MP's observed effects are exclusively placebo effects. It is possible that MP's therapeutic effects are nothing but placebo effects, but it cannot be denied that MDMA, in itself, promotes prosocial and other effects independent of expectancies. Octopuses and human evolution diverged about 700 million years ago from the flatworm, which has an oxytocin receptor system (Kobayashi et al., 2022). It is hard to believe that octopuses have expectancies linked to MDMA administration if they've never tried it before. A study have shown that MDMA causes prosocial behavior in octopuses, whom are normally a socially reclusive species (Edsinger & Dölen, 2018). Such prosocial behavior is also widely observed in, and reported by, human MDMA-users (Baylen & Rosenberg, 2006; Carlyle et al., 2019; Hysek et al., 2014) and in rodent studies (Curry et al., 2017). This means that expectancies in human clinical use are definitely not mediating the prosocial effects of MDMA, and that the neurochemical prosocial readiness is deeply rooted in evolutionary history. The prosocial effects induced by MDMA probably enhances therapeutic alliance, and consequently therapeutic outcome (Johansen & Krebs, 2009). However, these prosocial effects likely still moderate and interact with placebo effects in various ways. One pathway may be that the strong subjective effects caused by MDMA, affirm the participants' belief that they have received the full-dose of MDMA, which they expect to be a good thing. Another pathway might be that participants have their belief in MP's efficacy

strengthened, by virtue of the prosocial effects being experienced as pleasurable. Similar arguments can be made for the three other primary MDMA-promoted neurotransmitters that mediate other psychological effects e.g., serotonin's fear-reducing characteristics. It is not far-fetched to assume that some of these effects are actually beneficial to the therapeutic process, as already suggested in the introduction of this review. Admittedly, the claim that MDMA enhances therapeutic outcomes independent of expectancies, just remains a theoretically supported assumption from biological and neuropsychological levels of analysis, and by the testimonials of therapists and patients. Thus far, it seems without successfully blinding participants in the MP trials, there cannot be clear and direct empirical support for MP's true effectiveness.

4.2.d: The Double-blind's Double-bind

There is a paradoxical nature to the double-blind RCT design, in the context of psychotherapy- and pharmacotherapy-research. Mental health patients naturally seek treatment because they wish to feel better, and so research participants may guess what group they have been allocated to, not based on the taste or color of the pill, but based on their emotional improvement (Kirsch, 2009). Small true therapeutic effects may therefore unblind participants, to the degree of emotional improvement. This emotional improvement is presumably closely associated with the quantified outcome scores. Those who feel better, report better outcomes, and are more likely to guess that they have been allocated to the active treatment. If that line of reasoning is true, then clinical interventions with true therapeutic effects tend to have imperfect blinding success. The predicament here is that it may not be possible to create a perfectly convincing placebo-control for psychotherapies with true therapeutic effects. If a placebo-control were indeed perfect in terms of blinding participants, observed at 50% incorrect guesses of allocation, then it is likely due to a relatively weak or diffuse active treatment. As mentioned, some antidepressant trials, with small effect sizes, have reported imperfect blinding at around 40% incorrect guess rates (Lin et al., 2022), and the true rate may be even lower than that due to biased reporting (Scott et al., 2022). It begs the more general question as to whether the therapeutic effects of antidepressants are simply placebo effects. It has been suggested that the side effects of antidepressants are noticed by participants, who then assume they are on a trajectory toward healing, based on their expectancies (Moncrieff, 2008). In sum, the double-blind RCT design may be the gold standard in much of medicine, but it may need tweaking or a complete revamp to work as intended in psychiatric intervention studies, since the interventions are often easily discernable from placebos (Juul et al., 2021). Thus, imperfect blinding does not seem to be a unique feature of MP, but rather a widespread phenomenon in psychiatric research.

4.2.f: Findings for Social Anxiety in People with Autism

Social anxiety commonly occurs in people with autism (Spain et al., 2018). A previous systematic review, looking at interventions for social anxiety in people with autism, observed effect sizes ranging *d* = 0.10 to *d* = 1.48, with an average *d* = 0.60 (Wilson et al., 2019). In the present review, (Danforth et al., 2018) observed an effect size at *d* = 1.50, and moreover showed effect durability, which was not reported in the previous review. One might ask if the large effects observed was mainly due to the oxytocin releasing characteristic of MDMA. However, a similar previous study using oxytocin to treat social anxiety in people with autism, showed very poor durability of effects at *d* = 0.28, but with good acute effects (Watanabe et al., 2015). Keep in mind, the oxytocin study was purely a pharmacotherapy study. This may suggest that MDMA's non-oxytocin characteristics improves the therapeutic response compared to oxytocin alone. Additionally, the evidence may suggest that therapist support, during the acute effects of MDMA, facilitates the durability of the therapeutic effects. Regardless of mechanisms, the observed effects in (Danforth et al., 2018) were durable and were on par with the most effective treatments observed in the previous systematic review. However, there is a plethora of scales used to measure social skills and anxiety in people with autism, so the results may not be easily comparable.

4.3: Limitations

4.3.a: Included studies' limitations

The studies included in the present systematic review were a small set of studies with mostly small sample sizes, low blinding success, lack of expectancy measures, lack of LTFU control groups, and with only anxiety disorders as the target diagnosis. These factors limited the conclusion to the research question. Additionally, no formal risk of bias assessment was conducted. All the included studies were sponsored by MAPS, which may point to some bias, but at the same time to homogeneity of the treatment modality. A previous systematic review had already conducted a Cochrane Risk of Bias assessment of several of the included studies (Tedesco et al., 2021), and found that the studies provided moderate to high quality evidence. Furthermore, the discussion thus far already covers many of the factors relevant to risk of bias assessment, and covers most of the points included in a dedicated risk of bias assessment (Page et al., 2021).

4.3.b: Systematic Review Process Limitations

There was no review of the safety aspects of MP, specifically concerning the neurological and psychological risk profile of MDMA. It can be inferred that the safety profile of MDMA is at least safe enough for ethics committees to approve testing in humans. However, the potential adverse effects are important to estimate, since the risk vs. reward ratio is important for any treatment that bears any non-

trivial risks. MDMA's safety profile has been thoroughly covered in a previous systematic review (Tedesco et al., 2021), and the topic was deemed outside the scope of the present review.

While it was the case that only anxiety type disorder-studies strictly met the inclusion criteria, there were an alcohol abuse disorder (AUD) safety and tolerability study that included heavy drinking days as a secondary outcome measure (Sessa et al., 2021). It also happened to be the only near-eligible study that was not sponsored by MAPS. Furthermore, the AUD study showed some promising results, but was not included in the present systematic review, since the relevant outcome measure was not planned as a primary outcome. Similarly, there may have been other safety and tolerability studies that was not included in the present review.

Feasibility was also not formally included in the review process. Nevertheless, some inspirations were drawn from the reviewed literature. These inspirations are summarized here and carry some implications for future research priorities. One question is whether one supporting therapist is as effective as two therapists. Currently, the treatment protocol used in all the included studies, as developed by (MAPS, 2010), dictated two supporting therapists to be included in each therapy session. The trials that included three MDMA-dosed sessions had approximately a total of 40 therapy-hours pr. participant, including the preparatory-, dosed-, and integratory-therapy sessions. This means that the cost to potential clients in the future is 80 hours' worth of therapy, since the total hours are doubled by having two therapists. The costs of MDMA itself is negligible (Marseille et al., 2020), so the price of the therapy is significantly reduced if one therapist can manage the therapeutic process adequately. Even at 40-hours of therapy, this therapy is still a costly intervention, compared to treatment programs like Eye Movement Desensitization and Reprocessing therapy (EMDR), which usually does not take more than 10 sessions with a total of 15 therapy hours, and have also shown large effect sizes, as shown in one study (Van Der Kolk et al., 2007). However, this EMDR study showed a much lower effect for participants with childhood-onset PTSD, suggesting that MP may be a feasible treatment for that subgroup of patients i.e., complex PTSD. Direct comparison studies with LTFU may be needed to understand which other patient population subgroups MP may be feasible for. From a financial point of view, assuming that MP has great efficacy, it does not seem likely that its role will become that of a first-line treatment, but rather a reservist role for those most difficult-to-treat patients. In any case, the high amount of therapy hours involved in MP will be a significant factor in determining the feasibility of MP.

The present systematic review relied on average symptom reduction outcome scores and Cohen's d for the within-group effect sizes, whereas only Cohen's d was reported for the between-group effect sizes. The interpretability of the results was thus limited by not having explicit measures of response- and remission-rates, and relative symptom reduction scores for the between-group effect sizes. The response-

and remission-rates cannot be strictly inferred from the average symptom reduction scores. This limitation was chosen to balance the complexity of the data extraction- and analysis-process with the potential implications of the results.

The included studies were heterogeneous in terms of their chosen outcome scales, target disorders, symptom severity of the participants, MDMA dose levels for both the active groups and the control groups, and diagnosis lifetimes. Such heterogeneity may skew some of the implications drawn from the results, especially since the total number of included studies and included participants were relatively low, compared to systematic reviews of traditional psychotherapy and pharmacotherapy. Thus, for the present systematic review, to enable the inclusion of a diverse set of studies, it was decided a priori to avoid meta-analytical aggregation of the effect sizes, since this level of heterogeneity was expected. To enable future systematic reviewers to restrict the inclusion criteria, and the consequent level of heterogeneity between studies, the evidence base for MP has to grow significantly larger.

4.4: Implications

Since the pool of evidence for MP is still small and limited by low blinding success in particular, the evidence does not yet carry significant implications for society and mainstream therapy. MDMA remains an illicit drug in most countries, and so the first step is to produce an adequate evidence base that may eventually lead to medical legalization. If the effect sizes observed thus far are maintained as the evidence base grows in size and quality, there are potentially a significant number of patients, who are currently treatment-resistant that may be treated with MP. As such, some suggestions are presented as to how future research may deal with the blinding problem and what variables of interest should be prioritized.

4.4.a: Solving the Blinding Problem: Potential Study Designs

The normal solution to the blinding problem, is to create a placebo that is, in every way, indistinguishable from the active treatment (Barkham et al., 2021). This has indeed been the rationale behind using low-dose MDMA in control groups, also referred to as *active placebos* (Aday et al., 2022). Researchers thought that such a dose was low enough to be non-therapeutic, while simultaneously being high enough to convince participants and therapists of active treatment allocation. As observed in the included MP-studies, this active placebo strategy was more successful in blinding participants than using inert placebo pills, but it was not perfect and with great variance between studies (Mithoefer et al., 2018; Oehen et al., 2013; Ot'alora et al., 2018). Sometimes, new research paradigms are required when novel treatment modalities show up (Angus et al., 2015). Thus, some alternative strategies are now suggested for future research. The superordinate goal of these research design strategies is to allow differentiation of MP's true effectiveness from the placebo effects. This can be achieved in one of two general ways. Either perfect blinding, or expectancy-correction.

***Deception*.** One potential way to achieve perfect blinding, is to inform therapists and participants that they are partaking in an open-label study, and that all participants will receive 100mg MDMA during their MP-sessions (Muthukumaraswamy et al., 2021). However, in truth, 50% of the participants is given only 40mg of MDMA. If the 100mg MDMA group actually shows greater symptom reduction than the 40mg MDMA group, the evidence strongly suggests that MP's true therapeutic effects are enhanced by MDMA, since the placebo effects should be close to equal between the groups. The reason why the 40mg MDMA control should be employed, rather than just inert placebo, is because an inert placebo would probably draw too much suspicion from both participants and therapists. Ot'alora et al. (2018) used 40mg MDMA for their control group with relatively good success. The combined blinding ability of the low-dose MDMA and the deceiving nature of the design, will likely not raise suspicion in participants and therapists that they are getting less than 100mg MDMA. To further conceal the true objective of the trial, the decoy-objective could be almost any other research question e.g., the effects of one vs. two therapists pr. session. The biggest challenges with this design are probably ethical in nature (Haas et al., 2021). It is naturally considered unethical to withhold earnest treatments from people with mental disorders, and especially so in a deception study where participants are not even aware of the possibility. Alternatively, this type of study design could also be conducted under the flag of wanting to test MP for less severe, non-treatment resistant cases of anxiety or PTSD. This serves both as a decoy-reason for conducting the study and simultaneously making ethical approval more likely, given the less severe symptoms. If such an experiment established that MP in fact is effective, it would be a significant step toward MP's stamp of approval (Butlen-Ducuing et al., 2023).

***Positive Expectancy-adjusted Effect Sizes*.** A radical approach is to entirely do away with the standard of doing double-blinding. Rather, the expectancy levels of the participants are measured, and are subtracted from the within-group effect sizes by some standardized statistical method. One expectancy scale that could be implemented is the Credibility/Expectancy Questionnaire (Devilly & Borkovec, 2000), which have been observed to be predictive of psychiatric treatment outcomes (Webb et al., 2013). Admittedly, this is a radical approach, as it may require much of mainstream psychotherapy research to follow suit, to establish an evidence base for comparison outside of the traditional double-blinded RCT-paradigm. Mainstreaming this methodology would likely require decades of research, given that hundreds of different treatment modality- and diagnosis-combinations need to be tested. This method probably does not solve all problems associated with placebo, but it may still be a fruitful line of reasoning to pursue in psychotherapy research generally, and especially for MP.

***Expectancy-based Semi-randomization Between Two Active Treatments*.** Another way to avoid blinding, is to randomize participants between two different active interventions based on their expectancy

levels. At recruitment the participants are asked about their expectancy and/or preferred treatment between the two treatment modalities in question e.g., MP vs. EMDR. Participants are then selected and distributed, from the potential pool of recruits, between the two treatment groups to equalize the average expectancy levels. This ensures that there are near equal placebo effects between the treatment groups, despite all participants being entirely unblinded to what treatment they are receiving (Kirsch, 2009). A further variation of this study is to ensure that the participants in each intervention group, include participants who preferred the opposite treatment, as well as participants who preferred the allocated treatment. This way it can also be differentially shown how much more effective the treatment is for people with higher expectancy levels within and between each intervention group. Such findings may also carry important implications for real-world clinical practice. A large factor in what is truly recommendable to patients, may be the treatment they prefer, since that is where their expectancies are probably the highest.

***Practice-based Evidence in Addition to Evidence-based Practice*.** It is often argued that mental health professionals should follow the Scientist-Practitioner model and employ evidence-based interventions (Barkham et al., 2021). However, it seems that the practitioner part of that model is often left out from the evidence-base i.e., practice-based evidence seems somewhat neglected. Psychotherapy research is supposed to carry implications for real-world clinical practice. There are at least two good reasons for producing more practice-based evidence derived from real-world clinics or field studies. The first reason is that the results are arguably more generalizable to other similar routine real-world clinics. Evident from the prevalence of mental disorders, most mental health care practitioners will have clients who have not responded well to their treatments, and it is exactly those patients who are the most interesting to include in the MP studies. If MP shows great effectiveness in treating those patients, in field settings, that have not responded well to any of the standard care treatments, then that is surely strong supporting evidence in favor of MP's effectiveness and generalizability. Secondly, there are many variables associated with MP research that is interesting to experiment with, with the purpose of making MP more effective, safe, and feasible. Collaborating with real-world clinics significantly increases the sample size capacity for this field of research, and is almost essential to expand the evidence base. The total sample size of the included studies in this systematic review was (N = 264), after decades of research. The low amount of research may partly have something to do with the low profit-potential that MP carries for commercial entities, as MDMA cannot be patented, and all the MP research thus far have been philanthropically funded through MAPS. As such, it may be necessary for researchers to collaborate with state-sponsored clinics to ramp up the rate of research, to be able to cover all the relevant experimental variables and patient populations.

4.4.b: Experimental Variables To be Studied

Beyond the overall effectiveness of MP, a reading of the MP clinical studies inspires a host of ideas, as to what the most effective and practical approach for MP is. Each of those ideas gives birth to new research questions and hypotheses. Some of those questions and related variables are mentioned here.

What is the dose response curve for number of MP-sessions? To test hypotheses related to this question, one might simply conduct open-label studies with participants completing MP-therapy between one and five, or more, total MP-sessions. This can show whether MP's treatment effects cumulate beyond three MP-sessions, and whether there is diminishing returns toward the sessions at the end of treatment (Robinson et al., 2020).

Is one therapist as effective as two therapists? As mentioned, this is an interesting question primarily concerning the feasibility of MP. If one therapist can achieve the same, or close to the same, as two therapists, then the cost-efficiency is almost doubled (Marseille et al., 2020).

What are the most optimal treatment guidelines for MP? The protocol-guidelines for MP are axiomatic to some degree, since MP is still a novel treatment, and its primary experimental variable in all studies thus far has singularly been the dosage of MDMA. The guidelines have been based on some previous psychedelic research, but also on anecdotal evidence (MAPS, 2010). It would be interesting to test these treatment guidelines, like one would test the effectiveness between a psychodynamic therapy and a cognitive therapy, both augmented by MDMA. For example, nurturing touch is part of what is recommended in the MP treatment protocol. Another example is that participants are usually encouraged to get comfortable in a bed during the dosed sessions. A third example is eyeshades and music through headphones are encouraged to be utilized during the dosed sessions.

The first and foremost research priority should be in including patients with histories of treatment resistance, or patients with disorders that are usually difficult to treat, from a diverse set of mental disorders. From a pragmatic point of view, one can argue that if MP works well in the most desperate cases, questions as to why and how MP works are of secondary importance.

4.5: Conclusion of the Systematic Review

The large symptom reductions observed across the included studies in the present systematic review were promising. The large effects were observed in difficult-to-treat participants, that may well represent some of the most difficult-to-treat cases in real clinical settings. It is exactly those patients with treatment-resistance that desperately need more effective psychotherapy. However, the evidence base for MP was small compared to traditional psychotherapy and pharmacotherapy, and were limited to anxiety disorders only. Furthermore, most studies had low success in blinding their participants and therapists to

treatment allocation, suggesting that an unknown proportion of the observed therapeutic effects were due to enhanced placebo effects in the active treatment groups. Thus, it remains unclear how effective MP is beyond placebo effects, but more research is surely warranted. To produce more convincing evidence in the future, researchers may need to innovate study designs that reduce the uncertainty around the placebo effects. Additionally, to produce a much greater volume of research within a foreseeable future, and to ensure MP's generalizability beyond anxiety disorders and to routine practice, it may be necessary for researchers to collaborate with real-world clinics, increasing the currently slow rate of research.

References

Aday, J. S., Heifets, B. D., Pratscher, S. D., Bradley, E., Rosen, R., & Woolley, J. D. (2022). Great Expectations: recommendations for improving the methodological rigor of psychedelic clinical trials. *Psychopharmacology*, *239*(6), 1989-2010. https://doi.org/10.1007/s00213-022-06123-7

Albert, P. R., Benkelfat, C., & Descarries, L. (2012). The neurobiology of depression—revisiting the serotonin hypothesis. I. Cellular and molecular mechanisms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1601), 2378-2381. https://doi.org/10.1098/rstb.2012.0190

Amoroso, T. (2015). The Psychopharmacology of ±3,4 Methylenedioxymethamphetamine and its Role in the Treatment of Posttraumatic Stress Disorder. *Journal of psychoactive drugs*, *47*(5), 337-344. https://doi.org/10.1080/02791072.2015.1094156

Andrade, C. (2022). Psychotropic Drugs With Long Half-Lives. *The journal of clinical psychiatry*, *83*(4). https://doi.org/10.4088/jcp.22f14593

Angus, L., Watson, J. C., Elliott, R., Schneider, K., & Timulak, L. (2015). Humanistic psychotherapy research 1990-2015: From methodological innovation to evidence-supported treatment outcomes and beyond. *Psychotherapy Research*, *25*(3), 330-347. https://doi.org/10.1080/10503307.2014.989290

Bahji, A., Forsyth, A., Groll, D., & Hawken, E. R. (2019). Efficacy of 3,4-methylenedioxymethamphetamine (MDMA)-assisted psychotherapy for posttraumatic stress disorder: A systematic review and meta-analysis [Review]. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 109735.

Barkham, M., Lutz, W., Castonguay, L. G., Bergin, A. E., & Garfield, S. L. (2021). *Bergin and Garfield's Handbook of psychotherapy and behavior change*. Wiley.

Baylen, C. A., & Rosenberg, H. (2006). A review of the acute subjective effects of MDMA/ecstasy. *Addiction (Abingdon, England)*, *101*(7), 933-947. https://doi.org/10.1111/j.1360-0443.2006.01423.x

Bedi, G., Cotton, S. M., Guerin, A. A., & Jackson, H. J. (2022). MDMA-assisted psychotherapy for post-traumatic stress disorder: The devil is in the detail. *Aust N Z J Psychiatry*, 48674221127186. https://doi.org/10.1177/00048674221127186

Bell, M. L., Kenward, M. G., Fairclough, D. L., & Horton, N. J. (2013). Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *BMJ*, *346*(jan21 1), e8668-e8668. https://doi.org/10.1136/bmj.e8668

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

Bouso, J. C., Doblin, R., Farré, M., Alcázar, M. A., & Gómez-Jarabo, G. (2008). MDMA-assisted psychotherapy using low doses in a small sample of women with chronic posttraumatic stress disorder. *J Psychoactive Drugs*, *40*(3), 225-236. https://doi.org/10.1080/02791072.2008.10400637

Butlen-Ducuing, F., McCulloch, D. E. W., Haberkamp, M., Mattila, T., Bałkowiec-Iskra, E., Aislaitner, G., Balabanov, P., Lundberg, J., Stenbæk, D. S., Elferink, A., Knudsen, G. M., & Thirstrup, S. (2023). The therapeutic potential of psychedelics: the European regulatory perspective. *The Lancet (British edition)*, *401*(10378), 714-716. https://doi.org/10.1016/S0140-6736(23)00264-7

Carhart-Harris, R., & Nutt, D. (2017). Serotonin and brain function: a tale of two receptors. *Journal of Psychopharmacology*, *31*(9), 1091-1120. https://doi.org/10.1177/0269881117725915

Carlyle, M., Stevens, T., Fawaz, L., Marsh, B., Kosmider, S., & Morgan, C. J. A. (2019). Greater empathy in MDMA users. *Journal of psychopharmacology (Oxford)*, *33*(3), 295-304. https://doi.org/10.1177/0269881119826594

Carvalho, A. F., & McIntyre, R. S. (2015). *Treatment-Resistant Mood Disorders*. Oxford Psychiatry Library.

Chekroud, A. M., Foster, D., Zheutlin, A. B., Gerhard, D. M., Roy, B., Koutsouleris, N., Chandra, A., Esposti, M. D., Subramanyan, G., Gueorguieva, R., Paulus, M., & Krystal, J. H. (2018). Predicting Barriers to Treatment for Depression in a U.S. National Sample: A Cross-Sectional, Proof-of-Concept Study. *Psychiatric Services*, *69*(8), 927-934. https://doi.org/10.1176/appi.ps.201800094

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., Leucht, S., Ruhe, 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

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

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, A., & Furukawa, T. A. (2020). A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*, *19*(1), 92-107. https://doi.org/10.1002/wps.20701

Curry, D., Belkoff, A., & Howell, L. (2017). Prosocial effects and neurotoxicity of (-)-3,4-methylenedioxymethamphetamine in mice [Conference Abstract]. *Drug and Alcohol Dependence*, *171*, e50.

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 (Berl)*, *235*(11), 3137-3148. https://doi.org/10.1007/s00213-018-5010-9

De Crescenzo, F., D'Alò, G. L., Ostinelli, E. G., Ciabattini, M., Di Franco, V., Watanabe, N., Kurtulmus, A., Tomlinson, A., Mitrova, Z., Foti, F., Del Giovane, C., Quested, D. J., Cowen, P. J., Barbui, C., Amato, L., Efthimiou, O., & Cipriani, A. (2022). Comparative effects of pharmacological interventions for the acute and long-term management of insomnia disorder in adults: a systematic review and network meta-analysis. *The Lancet*, *400*(10347), 170-184. https://doi.org/10.1016/s0140-6736(22)00878-9

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of behavior therapy and experimental psychiatry*, *31*(2), 73-86. https://doi.org/10.1016/S0005-7916(00)00012-4

DHDA. (2021). *Burden of Sickness in Denmark*. Danish Health Data Authority.

Dragioti, E., Karathanos, V., Gerdle, B., & Evangelou, E. (2017). Does psychotherapy work? An umbrella review of meta-analyses of randomized controlled trials. *Acta psychiatrica Scandinavica*, *136*(3), 236-246. https://doi.org/10.1111/acps.12713

Edsinger, E., & Dölen, G. (2018). A Conserved Role for Serotonergic Neurotransmission in Mediating Social Behavior in Octopus. *Current Biology*, *28*(19), 3136-3142.e3134. https://doi.org/10.1016/j.cub.2018.07.061

Elton, C. (2019). *The Interview: MDMA-Therapy Expert Dr. Rick Doblin*.

FDA, F. D. A. (2017). *Breakthrough Therapy Designation*

Finniss, D. G. D., Kaptchuk, T. J., Miller, F. P., & Benedetti, F. P. (2010). Biological, clinical, and ethical advances of placebo effects. *The Lancet (British edition)*, *375*(9715), 686-695. https://doi.org/10.1016/S0140-6736(09)61706-2

Freudenmann, R. W., Öxler, F., & Bernschneider-Reif, S. (2006). The origin of MDMA (ecstasy) revisited: the true story reconstructed from the original documents. *Addiction*, *101*(9), 1241-1245. https://doi.org/10.1111/j.1360-0443.2006.01511.x

Green, A. R., Mechan, A. O., Elliott, J. M., O'Shea, E., & Colado, M. I. (2003). The Pharmacology and Clinical Pharmacology of 3,4-Methylenedioxymethamphetamine (MDMA, â€œEcstasyâ€). *Pharmacological reviews*, *55*(3), 463-508. https://doi.org/10.1124/pr.55.3.3

Gøtzsche, P. C. (2016). Antidepressants are addictive and increase the risk of relapse. *BMJ (Online)*, *352*, i574-i574. https://doi.org/10.1136/bmj.i574

Hinton, E., Steel, Z., Hilbrink, D., & Berle, D. (2022). Anger and predictors of drop-out from PTSD treatment of veterans and first responders. *Behavioural and cognitive psychotherapy*, *50*(2), 237-251. https://doi.org/10.1017/S1352465821000382

Huhn, M., Tardy, M., Spineli, L. M., Kissling, W., Förstl, H., Pitschel-Walz, G., Leucht, C., Samara, M., Dold, M., Davis, J. M., & Leucht, S. (2014). Efficacy of Pharmacotherapy and Psychotherapy for Adult Psychiatric Disorders. *JAMA Psychiatry*, *71*(6), 706. https://doi.org/10.1001/jamapsychiatry.2014.112

Hysek, C. M., Schmid, Y., Simmler, L. D., Domes, G., Heinrichs, M., Eisenegger, C., Preller, K. H., Quednow, B. B., & Liechti, M. E. (2014). MDMA enhances emotional empathy and prosocial behavior. *Social Cognitive and Affective Neuroscience*, *9*(11), 1645-1652. https://doi.org/10.1093/scan/nst161

Haas, J. W., Rief, W., & Doering, B. K. (2021). Open-Label Placebo Treatment: Outcome Expectations and General Acceptance in the Lay Population. *International Journal of Behavioral Medicine*, *28*(4), 444-454. https://doi.org/10.1007/s12529-020-09933-1

Illingworth, B., Lewis, D., Lambarth, A., Stocking, K., Duffy, J., Jelen, L., & Rucker, J. (2021). A comparison of mdma-assisted psychotherapy to non-assisted psychotherapy in treatment-resistant PTSD: A systematic review and meta-analysis [Conference Abstract]. *European Psychiatry*, *64(Supplement 1)*, S141-S142.

Jardim, A. V., Jardim, D. V., Chaves, B. R., Steglich, M., Ot'alora, G. M., Mithoefer, M. C., da Silveira, D. X., Tofoli, L. F., Ribeiro, S., Matthews, R., Doblin, R., & Schenberg, E. E. (1999). 3,4-methylenedioxymethamphetamine (MDMA)-assisted psychotherapy for victims of sexual abuse with severe post-traumatic stress disorder: an open label pilot study in Brazil. *Revista Brasileira de Psiquiatria*(pagination).

Johansen, P., & Krebs, T. (2009). How could MDMA (ecstasy) help anxiety disorders? A neurobiological rationale. *Journal of Psychopharmacology*, *23*(4), 389-391.

Juul, S., Gluud, C., Simonsen, S., Frandsen, F. W., Kirsch, I., & Jakobsen, J. C. (2021). Blinding in randomised clinical trials of psychological interventions: a retrospective study of published trial reports. *BMJ evidence-based medicine*, *26*(3), 109-109. https://doi.org/10.1136/bmjebm-2020-111407

Kaptchuk, T. J. (1998). Intentional Ignorance: A History of Blind Assessment and Placebo Controls in Medicine. *Bulletin of the history of medicine*, *72*(3), 389-433. https://doi.org/10.1353/bhm.1998.0159

Kaushik, A., Kostaki, E., & Kyriakopoulos, M. (2016). The Stigma of Mental Illness in Children & Adolescents: A Systematic Review. *Psychiatry research*, *243*, 469-494. https://doi.org/10.1016/j.psychres.2016.04.042

Keller, M. B., Lavori, P. W., Mueller, T. I., Endicott, J., Coryell, W., Hirschfeld, R. M., & Shea, T. (1992). Time to recovery, chronicity, and levels of psychopathology in major depression. A 5-year prospective follow-up of 431 subjects. *Arch Gen Psychiatry*, *49*(10), 809-816. https://doi.org/10.1001/archpsyc.1992.01820100053010

Khan, A., Redding, N., & Brown, W. A. (2008). The persistence of the placebo response in antidepressant clinical trials. *Journal of Psychiatric Research*, *42*(10), 791-796. https://doi.org/10.1016/j.jpsychires.2007.10.004

Kirsch, I. (2009). *The emperor's new drugs: exploding the antidepressant myth*. Bodley Head. https://go.exlibris.link/Xdxrrs6v

Kobayashi, A., Hamada, M., Yoshida, M.-a., Kobayashi, Y., Tsutsui, N., Sekiguchi, T., Matsukawa, Y., Maejima, S., Gingell, J. J., Sekiguchi, S., Hamamoto, A., Hay, D. L., Morris, J. F., Sakamoto, T., & Sakamoto, H. (2022). Vasopressin-oxytocin–type signaling is ancient and has a conserved water homeostasis role in euryhaline marine planarians. *Science Advances*, *8*(9), eabk0331. https://doi.org/doi:10.1126/sciadv.abk0331

Kong, W., Deng, H., Wan, J., Zhou, Y., Zhou, Y., Song, B., & Wang, X. (2020). Comparative Remission Rates and Tolerability of Drugs for Generalised Anxiety Disorder: A Systematic Review and Network Meta-analysis of Double-Blind Randomized Controlled Trials. *Front Pharmacol*, *11*, 580858. https://doi.org/10.3389/fphar.2020.580858

Leichsenring, F., Abbass, A., Hilsenroth, M. J., Leweke, F., Luyten, P., Keefe, J. R., Midgley, N., Rabung, S., Salzer, S., & Steinert, C. (2017). Biases in research: risk factors for non-replicability in psychotherapy and pharmacotherapy research. *Psychological Medicine*, *47*(6), 1000-1011. https://doi.org/10.1017/s003329171600324x

Leichsenring, F., Steinert, C., Rabung, S., & Ioannidis, J. P. A. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: an umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry*, *21*(1), 133-145. https://doi.org/10.1002/wps.20941

Leppink, J., O'Sullivan, P., & Winston, K. (2016). Effect size – large, medium, and small. *Perspectives on Medical Education*, *5*(6), 347-349. https://doi.org/10.1007/s40037-016-0308-y

Li, Y., Xin, Y., Lu, K., Du, W., & Guo, F. (2021). Will Smog Cause Mental Health Problems? Indication from a Microsurvey of 35 Major Cities in China. *International Journal of Environmental Research and Public Health*, *18*(23), 12388. https://doi.org/10.3390/ijerph182312388

Lin, Y.-H., Sahker, E., Shinohara, K., Horinouchi, N., Ito, M., Lelliott, M., Cipriani, A., Tomlinson, A., Baethge, C., & Furukawa, T. A. (2022). Assessment of blinding in randomized controlled trials of

antidepressants for depressive disorders 2000–2020: A systematic review and meta-analysis. *EClinicalMedicine*, *50*, 101505. https://doi.org/10.1016/j.eclinm.2022.101505

Locher, C., Koechlin, H., Gaab, J., & Gerger, H. (2019). The Other Side of the Coin: Nocebo Effects and Psychotherapy. *Frontiers in Psychiatry*, *10*, 555-555. https://doi.org/10.3389/fpsyt.2019.00555

MAPS. (2010). MDMA-Assisted Therapy Treatment Manual. In. Multidisciplinary Assocation for Psychedelic Studies.

Marseille, E., Kahn, J. G., Yazar-Klosinski, B., & Doblin, R. (2020). The cost-effectiveness of MDMA-assisted psychotherapy for the treatment of chronic, treatment-resistant PTSD. *PLoS One*, *15*(10), e0239997. https://doi.org/10.1371/journal.pone.0239997

Mavranezouli, I., Megnin-Viggars, O., Daly, C., Dias, S., Welton, N. J., Stockton, S., Bhutani, G., Grey, N., Leach, J., Greenberg, N., Katona, C., El-Leithy, S., & Pilling, S. (2020). Psychological treatments for post-traumatic stress disorder in adults: a network meta-analysis. *Psychological Medicine*, *50*(4), 542-555. https://doi.org/10.1017/s0033291720000070

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. *Lancet Psychiatry*, *5*(6), 486-497. https://doi.org/10.1016/s2215-0366(18)30135-4

Mithoefer, M. C., Wagner, M. T., Mithoefer, A. T., Jerome, L., & Doblin, R. (2011). 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. *J Psychopharmacol*, *25*(4), 439-452. https://doi.org/10.1177/0269881110378371

Mithoefer, M. C., Wagner, M. T., Mithoefer, A. T., Jerome, L., Martin, S. F., Yazar-Klosinski, B., Michel, Y., Brewerton, T. D., & Doblin, R. (2013). Durability of improvement in post-traumatic stress disorder symptoms and absence of harmful effects or drug dependency after 3,4-methylenedioxymethamphetamine-assisted psychotherapy: a prospective long-term follow-up study. *J Psychopharmacol*, *27*(1), 28-39. https://doi.org/10.1177/0269881112456611

Mitsikostas, D., & Benedetti, F. (2019). *Placebos and Nocebos in headaches*. Springer International Publishing.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Group, P., & and the, P. G. (2009). Reprint—Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Physical therapy*, *89*(9), 873-880. https://doi.org/10.1093/ptj/89.9.873

Moncrieff, J. (2008). *The myth of the chemical cure: a critique of psychiatric drug treatment*. Palgrave Macmillan. https://go.exlibris.link/gg69KHxw

Moncrieff, J. (2018). What does the latest meta-analysis really tell us about antidepressants? *Epidemiology and Psychiatric Sciences*, *27*(5), 430-432. https://doi.org/10.1017/s2045796018000240

Moncrieff, J., Cooper, R. E., Stockmann, T., Amendola, S., Hengartner, M. P., & Horowitz, M. A. (2022). The serotonin theory of depression: a systematic umbrella review of the evidence. *Molecular psychiatry*. https://doi.org/10.1038/s41380-022-01661-0

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. *Eur J Psychotraumatol*, *11*(1), 1840123. https://doi.org/10.1080/20008198.2020.1840123

Murphy, D., & Smith, K. V. (2018). Treatment Efficacy for Veterans With Posttraumatic Stress Disorder: Latent Class Trajectories of Treatment Response and Their Predictors. *Journal of traumatic stress*, *31*(5), 753-763. https://doi.org/10.1002/jts.22333

Muthukumaraswamy, S. D., Forsyth, A., & Lumley, T. (2021). Blinding and expectancy confounds in psychedelic randomized controlled trials. *Expert Review of Clinical Pharmacology*, *14*(9), 1133-1152. https://doi.org/10.1080/17512433.2021.1933434

Nuñez, N. A., Joseph, B., Pahwa, M., Kumar, R., Resendez, M. G., Prokop, L. J., Veldic, M., Seshadri, A., Biernacka, J. M., Frye, M. A., Wang, Z., & Singh, B. (2022). Augmentation strategies for treatment resistant major depression: A systematic review and network meta-analysis. *J Affect Disord*, *302*, 385-400. https://doi.org/10.1016/j.jad.2021.12.134

Nøhr, A. K., Eriksson, H., Hobart, M., Moltke, I., Buller, R., Albrechtsen, A., & Lindgreen, S. (2021). Predictors and trajectories of treatment response to SSRIs in patients suffering from PTSD. *Psychiatry research*, *301*, 113964-113964. https://doi.org/10.1016/j.psychres.2021.113964

O'Reardon, J. P., & Amsterdam, J. D. (2001). Overview of treatment-resistant depression and its management. In *Treatment-Resistant Mood Disorders* (pp. 30-46). Cambridge University Press. https://doi.org/10.1017/CBO9780511666421.003

OECD. (2018). *Health at a Glance: Europe 2018*. Organisation for Economic Co-operation and Development. https://doi.org/10.1787/health_glance_eur-2018-en

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). *J Psychopharmacol*, *27*(1), 40-52. https://doi.org/10.1177/0269881112464827

Oliva, V., Lippi, M., Paci, R., Del Fabro, L., Delvecchio, G., Brambilla, P., De Ronchi, D., Fanelli, G., & Serretti, A. (2021). Gastrointestinal side effects associated with antidepressant treatments in patients with major depressive disorder: A systematic review and meta-analysis. *Prog Neuropsychopharmacol Biol Psychiatry*, *109*, 110266. https://doi.org/10.1016/j.pnpbp.2021.110266

Ot'alora, G., Grigsby, J., Poulter, B., Van Derveer, J. W., III, 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.

Page, M. J., Moher, D., 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., . . . McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, n160. https://doi.org/10.1136/bmj.n160

Papakostas, G. I., Jackson, W. C., Rafeyan, R., & Trivedi, M. H. (2020). Inadequate Response to Antidepressant Treatment in Major Depressive Disorder. *The journal of clinical psychiatry*, *81*(3). https://doi.org/10.4088/jcp.ot19037com5

Papaseit, E., Pérez-Mañá, C., Mateus, J.-A., Pujadas, M., Fonseca, F., Torrens, M., Olesti, E., De La Torre, R., & Farré, M. (2016). Human Pharmacology of Mephedrone in Comparison with MDMA. *Neuropsychopharmacology*, *41*(11), 2704-2713. https://doi.org/10.1038/npp.2016.75

Robinson, L., Delgadillo, J., & Kellett, S. (2020). The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research*, *30*(1), 79-96. https://doi.org/10.1080/10503307.2019.1566676

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., Niederehe, G., Thase, M. E., Lavori, P. W., Lebowitz, B. D., McGrath, P. J., Rosenbaum, J. F., Sackeim, H. A., Kupfer, D. J., Luther, J., & Fava, M. (2006). Acute and Longer-Term Outcomes in Depressed Outpatients

Requiring One or Several Treatment Steps: A STAR*D Report. *American Journal of Psychiatry*, *163*(11), 1905-1917. https://doi.org/10.1176/ajp.2006.163.11.1905

Scott, A. J., Sharpe, L., & Colagiuri, B. (2022). A systematic review and meta-analysis of the success of blinding in antidepressant RCTs. *Psychiatry research*, *307*, 114297-114297. https://doi.org/10.1016/j.psychres.2021.114297

Sessa, B., Higbed, L., O'Brien, S., Durant, C., Sakal, C., Titheradge, D., Williams, T. M., Rose-Morris, A., Brew-Girard, E., Burrows, S., Wiseman, C., Wilson, S., Rickard, J., & Nutt, D. J. (2021). First study of safety and tolerability of 3,4-methylenedioxymethamphetamine-assisted psychotherapy in patients with alcohol use disorder. *J Psychopharmacol*, *35*(4), 375-383. https://doi.org/10.1177/0269881121991792

Shulgin, A., & Shulgin, A. (1991). *PIHKAL: A Chemical Love Story*.

Smith, K. W., Sicignano, D. J., Hernandez, A. V., & White, C. M. (2022). MDMA-Assisted Psychotherapy for Treatment of Posttraumatic Stress Disorder: A Systematic Review With Meta-Analysis. *The Journal of Clinical Pharmacology*, *62*(4), 463-471. https://doi.org/10.1002/jcph.1995

Soyka, M. (2017). Treatment of Benzodiazepine Dependence. *The New England Journal of Medicine*, *376*(12), 1147-1157. https://doi.org/https://doi.org/10.1056/NEJMra1611832

Spain, D., Sin, J., Linder, K. B., McMahon, J., & Happé, F. (2018). Social anxiety in autism spectrum disorder: A systematic review. *Research in autism spectrum disorders*, *52*, 51-68. https://doi.org/10.1016/j.rasd.2018.04.007

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. *Int J Epidemiol*, *43*(2), 476-493. https://doi.org/10.1093/ije/dyu038

Swift, J. K., & Greenberg, R. P. (2014). A Treatment by Disorder Meta-Analysis of Dropout From Psychotherapy. *Journal of psychotherapy integration*, *24*(3), 193-207. https://doi.org/10.1037/a0037512

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

Trautmann, S., Rehm, J., & Wittchen, H. U. (2016). The economic costs of mental disorders. *EMBO reports*, *17*(9), 1245-1249. https://doi.org/10.15252/embr.201642951

Van Der Kolk, B. A., Spinazzola, J., Blaustein, M. E., Hopper, J. W., Hopper, E. K., Korn, D. L., & Simpson, W. B. (2007). A randomized clinical trial of eye movement desensitization and reprocessing (EMDR),

fluoxetine, and pill placebo in the treatment of posttraumatic stress disorder : Treatment effects and long-term maintenance. *The journal of clinical psychiatry*, *68*(1), 37-46. https://doi.org/10.4088/JCP.v68n0105

Vizeli, P., & Liechti, M. E. (2018). Oxytocin receptor gene variations and socio-emotional effects of MDMA: A pooled analysis of controlled studies in healthy subjects. *PLoS One*, *13*(6), e0199384-e0199384. https://doi.org/10.1371/journal.pone.0199384

Voineskos, D., Daskalakis, Z. J., & Blumberger, D. M. (2020). <p>Management of Treatment-Resistant Depression: Challenges and Strategies</p&gt. *Neuropsychiatric Disease and Treatment*, *Volume 16*, 221-234. https://doi.org/10.2147/ndt.s198774

Wampold, B., Mondin, G., Moody, M., Stich, F., Benson, K., & Ahn, H.-n. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes". *Psychological Bulletin - PSYCHOL BULL*, *122*, 203-215. https://doi.org/10.1037//0033-2909.122.3.203

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, *14*(3), 270-277. https://doi.org/10.1002/wps.20238

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*, 002216782110236. https://doi.org/10.1177/00221678211023663

Watanabe, T., Kuroda, M., Kuwabara, H., Aoki, Y., Iwashiro, N., Tatsunobu, N., Takao, H., Nippashi, Y., Kawakubo, Y., Kunimatsu, A., Kasai, K., & Yamasue, H. (2015). Clinical and neural effects of six-week administration of oxytocin on core symptoms of autism. *Brain (London, England : 1878)*, *138*(Pt 11), 3400-3412. https://doi.org/10.1093/brain/awv249

Webb, C. A., Kertz, S. J., Bigda-Peyton, J. S., & Björgvinsson, T. (2013). The role of pretreatment outcome expectancies and cognitive–behavioral skills in symptom improvement in an acute psychiatric setting. *Journal of Affective Disorders*, *149*(1-3), 375-382. https://doi.org/10.1016/j.jad.2013.02.016

Whitaker, R. B. (2007). Reality Check: What Science Has to Tell Us about Psychiatric Drugs and Their Long-Term Effects. *Journal of college student psychotherapy*, *21*(3-4), 97-123. https://doi.org/10.1300/J035v21n03_05

Wilson, K. P., Steinbrenner, J. R., Kalandadze, T., & Handler, L. (2019). Interventions Targeting Expressive Communication in Adults with Autism Spectrum Disorders: A Systematic Review. *Journal of speech, language, and hearing research*, *62*(6), 1959-1978. https://doi.org/10.1044/2018_JSLHR-L-18-0219

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., Van Os, J., Preisig, M., Salvador-Carulla, L., Simon, R., & Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, *21*(9), 655-679. https://doi.org/10.1016/j.euroneuro.2011.07.018

Wolde, A. K. (2022). The Cause of Mental Health Problem Among Urban Residents During the COVID-19 Outbreak: Implication for information Communication. *Illness, crisis, and loss*, *30*(3), 489-498. https://doi.org/10.1177/1054137320985291

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. *Sci Rep*, *10*(1), 20442. https://doi.org/10.1038/s41598-020-75706-1

Zhou, X., Teng, T., Zhang, Y., Del Giovane, C., Furukawa, T. A., Weisz, J. R., Li, X., Cuijpers, P., Coghill, D., Xiang, Y., Hetrick, S. E., Leucht, S., Qin, M., Barth, J., Ravindran, A. V., Yang, L., Curry, J., Fan, L., Silva, S. G., . . . Xie, P. (2020). Comparative efficacy and acceptability of antidepressants, psychotherapies, and their combination for acute treatment of children and adolescents with depressive disorder: a systematic review and network meta-analysis. *The Lancet Psychiatry*, *7*(7), 581-601. https://doi.org/10.1016/s2215-0366(20)30137-1