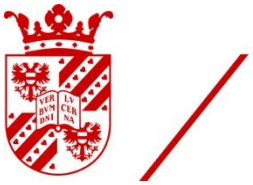


rijksuniversiteit groningen

Randomised Controlled Trials in the Field of Psychotic Disorders: What do RCTs Look Like?

Beau Meenhorst

Master Thesis – Klinische Psychologie

S5085667

August 15th 2023

Department of Psychology

University of Groningen

Examiner/Daily

supervisor: Drs. Nina Schwarzbach

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned.

Acknowledgements

I cannot begin to express my thanks to drs. Nina Schwarzbach. She gave me valuable suggestions, constructive criticism and had helpful contributions and profound belief in my work. I would also like to extend my gratitude to prof. dr. Henk Kiers and dr. Rink Hoekstra, who extended assistance with their extensive knowledge and helpful contributions, and to prof. dr. Marieke Pijnenborg for the practical suggestions and constructive criticism.

Abstract

Mental health problems are reported all over the world. In order to offer effective mental healthcare, psychological interventions and treatments should be tested to ensure effectiveness and generalizability. RCTs can potentially be seen as the current gold standard in clinical research, but there are doubts about its 'highest-quality' status, such as doubts about standardisation, generalizability and methodological/statistical choices made in RCTs. Therefore, a systematic literature review on RCTs on psychological interventions for psychotic disorders was conducted to get an overview of what RCTs look like in terms of methodology and analysis. Articles published between 2021 and 2023 on Pubmed and PsycINFO were screened on abstract/title and full-text. Simple random sampling was used for the final selection (n = 8). The methodological and statistical choices, results and conclusions were summarised and a quality review conducted. I found variation in the methodology of the RCTs because the design is a contextual factor depending on the aim of the study, which seems contradictory to standardisation of research. RCTs look alike in terms of participant inclusion, trial designs, group design and statistical analyses. All studies reported a positive trial. The quality varied and most have poor reporting and poor or missing justification of choices. Limitations are the search string, the small scope and the subjective quality review. Follow-up research may benefit from revising the eligibility criteria, broadening the scope and adding a critical reflection on the implications of choices and the quality of RCTs. For daily clinical practice it is advised to stay critical and combine information on the efficacy of treatments from multiple sources of evidence.

Keywords: randomised controlled trials, psychotic disorders, methodology, quality, psychological interventions

Contents

Randomised Controlled Trials in the Field of Psychotic Disorders: What do RCTs Look Like?

Mental health problems are reported all around the world (GBD 2019 Mental Disorders Collaborators, 2022). As many as 48% of Dutch adults have had a mental illness at least once at some point in their lives (ten Have et al., 2023). In the last 12 months, as many as 26% of Dutch adults have had a mental illness and 22% of Dutch adults sought help in mental healthcare for their problems (ten Have et al., 2023). These numbers show the immense impact of mental health problems on society and consequently the importance of adequate and effective mental healthcare. In order to offer adequate and effective mental healthcare to everyone, psychological interventions and treatments should be tested to ensure effectiveness and broad generalizability (Guidi et al., 2018).

The testing of interventions and treatments to ensure effectiveness and broad generalisation has always been a core feature in the field of Clinical Psychology and is part of the basis of *evidence-based practice* (EBP) (American Psychiatric Association, 2006). EBP, sometimes also referred to as evidence-based mental health (EMBH), is "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences'' (American Psychiatric Association, 2006, p. 273; see also Lillienfeld et al., 2013). From this perspective, the use of EBP and the testing of interventions is important because rigorous clinical research has shown that there are multiple interventions that can effectively reduce symptoms of mental disorders, eliminate risk factors and eliminate risk of onset for mental disorders (Becker, 2009).

EBP can be based on evidence based nomothetic research, which means using guidelines and supported treatments in the studies to create mental healthcare focused on standardisation of treatment, or on idiographic research, which means combining supported treatments, clinical expertise and patient's needs in the studies to create mental health care focused on the individual, or both (Hunsley, 2015; Lillienfeld et al., 2013). The nomothetic

assumptions are translated to guidelines and protocols, which are commonly used instruments in daily practice (Hunsley, 2015; Lillienfeld et al., 2013).

However, recent insights show us that the most effective intervention for a specific disorder may not always be the best treatment option for an individual (Stronks et al., 2013). Patient-bound factors such as personal characteristics, culture and preferences play a part in determining the effectiveness of a specific intervention on the specific psychological complaints of the individual and might be overlooked (Piccirillo & Rodebaugh, 2019; Stronks et al., 2013).

According to clinical practice guidelines, such as the NICE guidelines for mental health care (GBD 2019 Mental Disorders Collaborators, 2022), no intervention would be considered empirically supported unless at least one *randomised controlled trial* (RCT) has reported the effectiveness of the intervention (Reddy et al., 202l; see also Zismani et al., 2021). One example of an effective and empirically supported therapy is Cognitive Behavioral Therapy (CBT) which is used to treat an array of mental disorders (Hayes & Hofmann, 2017; Shafran et al., 2009). Most CBT interventions are considered EBP because they have been shown effective on different populations and problems (Hayes & Hofmann, 2017; Reddy et al., 2020). An RCT can potentially be seen as the current gold standard in clinical research (Reddy et al., 2020; Tackett et al., 2019). An RCT is an experimental study design in which participants are randomly assigned to either one of two conditions; one experimental group, which receives the treatment that the study aims to test, and one control group, which receives either *treatment as usual* (TAU), no treatment at all or another treatment that is supposed to be tested. Afterwards the groups will be compared to see what treatment was more effective (Akobeng, 2005). Given that it is hard to blind a participant or clinician to a certain condition, RCTs are sometimes single- or rater-blinded (Karanicolas et al., 2010). In essence the rigid methodology of an RCT creates a set research design, which is

also what makes RCTs a popular design for conducting research and is the main selling point for RCTs (Ivarsson & Andersen, 2016). In addition, CONSORT provides a framework for reporting RCTs, which standardised reporting on all elements of an RCT (Altman et al., 2001). Furthermore, it is believed that RCTs have the strongest evidence for causality because treatment conditions are randomly assigned to participants, which leaves less room for error or treatment effect (Zabor et al., 2020).

Even though an RCT is considered to be the highest quality of research, there are some doubts about the 'highest-quality' status that an RCT receives (Carey & Stiles, 2016; see also Ivarsson & Andersen, 2016; Krauss, 2018). One of the doubts concerning RCTs is in regard to the generalizability of RCTs in daily practice (Carey & Stiles, 2016). RCTs are built around the idea of a controlled circumstance, where there is no influence of confounding variables, so that careful causal claims can be made regarding the intervention (Cartwright, 2011; Deaton & Cartwright, 2018). The rigid methodology and controlling for variables make that the intervention will be effective in yielding the desired outcome somewhere, in specific conditions, on some individuals of the population (Cartwright, 2011; Deaton & Cartwright, 2018). For daily clinical practice however, it is important to know whether the outcome is generalizable (i.e. applicable to the patients) in more complex and sub-optimal circumstances (Carey & Stiles, 2016). Some argue that this makes RCTs no better than any other design in generalising knowledge about the efficacy and causality of interventions (Cartwright, 2011; Deaton & Cartwright, 2018).

Additionally, there are multiple ways to set up the methodology of an RCT, which can also affect the results. Even though the RCTs and CONSORT are both designed to standardise, usage of rigid methodology and CONSORT will not guarantee standardisation. In reality, some RCTs are not standardised at all because some elements cannot be standardised. For example, in one RCT the control group may be receiving treatment whereas

in the other RCT, the control group may be receiving no treatment at all, which can impact the effect sizes found in the two RCTs and thus influence the outcomes of the study. Other methodological choices that can be made involve the duration of the treatment and the amount of measuring points to compare results, which can impact the outcome and thus the conclusion and recommendations following the results of the RCT, which can lead to the implementation of interventions that may not be effective long-term.

Furthermore, not every RCT uses the same statistical analyses, which may lead to different interpretations based on the same data (Hiebert et al., 2006). There are multiple statistical analyses that can be used to analyse the obtained data, such as a regression analysis, an ANOVA, or a multilevel analysis. Some RCTs also take into account confounding factors, which will result in the use of a moderation or mediation analysis instead of an ANOVA or regression. Sadly, the choice for one or more of these specific statistical analyses is hardly explained in depth. This can lead to confusion, difficulty interpreting the results and difficulty understanding the shortcomings of an article, especially if your understanding of statistics is limited. Furthermore, even when knowledgeable, not everyone has the ability, education or time to critically reflect on all choices that are made in an RCT and the outcomes of an RCT. The limited understanding of RCT choices and their implications are uncritically accepted for example by therapists (Cartwright, 2011).

Another important problem is the *file drawer effect*, a widely reported issue in science in general, but also in RCTs (Lemaire et al., 2022). Multiple RCTs get published every year in the field of psychology, but trials that report negative or neutral results are less likely to be published, meaning that positive results are most likely overrepresented, especially positive results from small trials, and most articles report false positives (Franco et al., 2014). These false positive RCTs indicate that an effect is present when the effect is not in fact present. In other words, they wrongly report a positive effect, which does not represent the truth (Franco

et al., 2014). There are various reasons for the file drawer effect (Lemaire et al., 2022). Scientific journals are interested in RCTs with significant results, which can lead to publication bias; neutral results are not published because they can be contradictory or not ground-breaking enough (Lemaire et al., 2022). Also, it can feel like one has failed when one has to report that the hypothesised results of the RCT are not found, thus making researchers hesitant to submit and publish their study (Lemaire et al., 2022).

Given that not all RCTs follow the same methodology and statistical analysis methods and that doubts arise regarding the credibility of RCTs if their main selling point is standardisation, one may wonder if RCTs should be the gold standard of research (Ivarsson & Andersen, 2016). To answer this question, I will give an overview of the current research standard and I will look critically at what an RCT looks like and if a 'standard RCT' exists at all. This has led to the following question: "What do randomised controlled trials of psychological interventions for psychotic disorders look like in terms of methodology and analysis?''. I will look at the (a) methodological and (b) statistical choices as well as (c) the results and interpretation of the results and (d) the quality of the different methodological and statistical aspects of the RCTs. The methodological choices include relevant information about the participants, sample size, inclusion and exclusion criteria, randomization, blinding, the groups, trial design, instruments, the intervention, control conditions, duration of the therapy, measuring points and outcome variable. The statistical choices include the statistical analysis chosen, assumptions of the chosen analytical method, outliers and removed participants. The results and interpretation include the influence of the trial, effect size, clinical relevance and the conclusion.

More specifically, I will critically examine the decisions that are taken in RCTs and summarise methodological and statistical choices. Differences in methodology and data-analysis are expected, for example differences in sample size, trial design, dealing with

assumptions and violations and the statistical analysis used to examine the data. Because all RCTs are using a different methodology and statistical analysis, differences in quality may be observed between the RCTs included, which is why a quality assessment will be conducted.

Methodology

Design

For this master's thesis research, an explorative systematic literature review will be conducted. The focus will be on RCTs for psychological interventions for psychotic disorders, with the articles published from 2021 to 2023 to get an overview of the current research standard. It is more feasible to create a framework for RCTs for one specific disorder; this ensures that the amount of available literature is not overwhelming and it makes it easier to compare between trials. Out of all eligible papers, a random selection of a maximum of 20 papers will be made by the master student for the data extraction. Simple random sampling will be used for the selection of articles for the final review. The methodology, statistical analysis and results of these articles will be quantified, and a quality assessment conducted.

The databases PsycINFO and PubMed were used to find relevant articles. Both databases are well-established databases and contain relevant articles in the field of Clinical Psychology.

Study selection process

Search strategy

For this review the focus will be on psychological interventions only. This means medication and other pharmacological interventions will be excluded, given that the main focus is on psychological interventions (e.g. CBT). To ensure most of the relevant articles regarding psychological interventions in the field of psychotic disorders were found, the following search string that was used: ("psychotic disorder*" OR schizophren* OR

"schizoaffective disorder*") AND (RCT OR randomised controlled trial OR randomized controlled trial) NOT (pharmaco* AND medic* AND psychiatr*). All articles emerging from this search will be screened but only published articles written in the English language will be included in the screening, and thus the review. Furthermore, only relevant articles published between January 2021 and April 2023 will be included in the screening and thus the review.

Screening

All literature that was found on PsycINFO and PubMed was uploaded in Rayyan, a web-tool for the screening and selection of papers in a systematic review (Ouzzani et al., 2016). After uploading all articles to Rayyan, a duplicate screening was conducted, and any duplicates were inspected and deleted if necessary. A thorough screening was conducted to assess the relevance of the articles found, which means that the literature was screened twice. The aim was to include around 500 articles in the first screening; in the first screening all titles and abstracts were screened for relevance by both the daily supervisor and the master student. If the title and/or abstract indicated that the article had something to do with RCTs and psychological interventions for psychotic disorders, then the abstract was screened using the inclusion criteria.

All articles that met our inclusion criteria in the title and abstract screening were downloaded and the full text was screened a second time by both the daily supervisor and master student. Interventions that focused solely on drug interventions and virtual reality (VR) were excluded. The daily supervisor and the student screened the articles independently and confirmed their final selection to make sure all relevant articles were included in the review. Any discrepancies were discussed and resolved.

Eligibility criteria

Inclusion criteria

Types of studies.

- The study must use an RCT; which regards random selection of groups, blinding of subjects, and an experimental and control condition.

Focus.

- Psychotic disorders; schizophrenia, schizophrenia spectrum disorder (SSD), schizoaffective disorder, brief psychotic disorder, delusion disorder and schizophreniform disorder.

Participants.

- Participants must be between the ages of 18 and 65.[1]
- Participants must have a diagnosis for a DSM-IV (American Psychiatric Association, 1994), DSM-IV-TR (American Psychiatric Association, 2000), DSM-5 (American Psychiatric Association, 2013), ICD-10 (World Health Organization, 1993) or ICD-11 (World Health Organization, 2022) defined psychotic disorder.

Intervention.

- It must be a psychological intervention, which uses talking interventions to introduce change in psychological symptoms and treat disorders; patients are allowed to use medication, transcranial magnetic stimulation (TMS) or movement as add-on, as long as the RCT's main topic is the psychological intervention. The sessions must be in person, either in a group or individually, but digitally supported interventions are allowed. The involvement of patients' families in the treatment is allowed, as long as the involvement leads to improvement for the patient only.

[1] The onset of psychotic disorders usually begins in the late teen years and adolescence (Kessler et al., 2007). Furthermore, individuals above the age of 65 years are more likely to develop psychotic symptoms as a result of dementia and an increased vulnerability to delirium (Brendel et al., 2005)

Exclusion criteria

Focus.

- ➢ The psychotic disorder should not be secondary to another disease.[2]
- ➢ The psychotic disorder must not arise from substance use, somatic disorders, or other mental disorders.

Intervention.

- ➢ Virtual reality interventions will be excluded.[3]
- ➢ Studies that use medication as a primary treatment will be excluded.

Data extraction

Eight random articles, drawn from the included articles, were assessed to get an overview of the different components of an RCT, which was used to set up a data-extraction table. Its capability to extract all relevant information was tested using three other random articles. Any missing information and discrepancies in the extraction table were brought up for discussion and the table was updated if needed. Additionally, after testing the extraction table on three articles and talking to experts in statistics, a consensus was reached about the content of the table.

The data-extraction table, see Table 1 in Appendix A, contains information about three different subtopics of the review, with the first one being the *methodology of the study*. This includes relevant information about for example the participants, sample size, trial design and outcome variable. The second subtopic is the *statistical analysis*, which includes information regarding the statistical analysis of the article, assumptions of the chosen

[2] Comorbidity may influence the expression of the symptoms and therefore the effectiveness of the intervention, which makes it difficult to compare to other RCTs where symptoms are to be explained by a psychotic disorder only.

[3] VR interventions are an addition to psychological interventions and therefore more standardisable than psychological interventions.

analytical method, outliers and removed participants. The third and final subtopic is the *result and conclusion*, which includes information about the effect size and clinical relevance.

Data-analysis plan

Summarising the methodology and statistical choices

All remaining articles were analysed. The variables of interest were (a) the methodology of the study, which included for example the outcome variable, the instruments used, the groups and the amount of measurement points, (b) the statistical analysis used, which also included assumptions, information about outliers and missing data, and (c) the results and conclusion. Then all methodological choices, statistical choices, results and conclusions were summarised and similarities and differences between the RCTs were discussed.

All data extraction was done by the master student. Before extracting the data from the randomly selected articles using simple random selection, both the daily supervisor and the master student used the extraction table to extract all relevant information from three articles for consensus about the content of the extraction table. Discrepancies in information were discussed with the supervisor. Missing information and other peculiarities were also discussed and resolved, after which the table was updated. When consensus about the extraction table was reached, the data from all articles was extracted by the master student. All questions regarding the final data extraction were discussed with the daily supervisor to resolve any issues.

Quality review

Every article that was included in the final review was subjected to a thorough and critical assessment of prespecified choices that were made in methodology and statistical analysis, to address the observed differences in RCTs. As a matter of fact, the observed differences may say something about the overall quality of RCTs in the field of psychotic

disorders. Factors that were of importance for this critical review were: (a) the methodology needs to be clearly defined, for example every choice that has been made must be based on reasonable argumentation and (b) every choice that has been made regarding the statistical analysis and the refining and usage of the data needs to be based on reasonable argumentation.

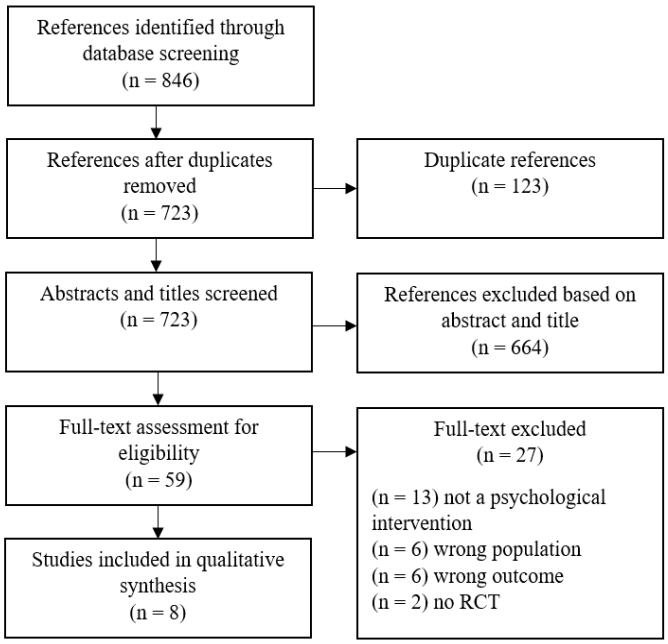
Results

Search

The search yielded a total of 846 articles, of which 123 were duplicates. After removing all duplicate articles, 723 articles were screened on title and abstract. After screening, 664 articles were excluded and 59 articles were deemed eligible for full-text screening. After full-text screening 32 articles were deemed eligible for inclusion in the review, which means 27 articles were excluded, as seen in the PRISMA flowchart in Figure 1. Eight randomly chosen articles were included in the final review. A detailed overview of the analysis can be found in Appendix A.

Figure 1

PRISMA flow chart



Summarising the methodology and statistical choices

Methodology of RCTs

Article 1, the study conducted by Hon-wai Chu et al. (2021), tested the effect of Positive Psychotherapy for Psychosis (PPP) on well-being and psychiatric symptomatology. Article 2, the study conducted by Böge et al. (2021), tested the effect of Mindfulness Based Group Therapy (MBGT) on mindfulness and psychiatric symptoms. Article 3, the study conducted by Zonp & Bilgin (2021), tested the effect of Metacognitive training (MCT) on social cognition. Article 4, the study conducted by Solar et al. (2022), tested the effect of a referral to and an appointment with an Acute Inpatient Unit (AIU) clinical psychologist on engagement in community therapy and auditory hallucinations and associated distress. Article 5, the study conducted by Longden et al. (2022), tested the effect of a Talking with Voices trial (TwV) on the distress accompanying auditory hallucinations. Article 6, the study conducted by Halverson et al. (2021), tested the effect of Integrated-Coping Awareness Therapy (I-CAT) on dysfunctional stress reactions and stress reactivity. Article 7, the study conducted by Weijers et al. (2021), tested the effect of Mentalization Based treatment for psychotic disorder (MBTp) on social functioning. Article 8, the study conducted by Cella et al. (2023), tested the effect of Group Training for Social skills in Psychosis (GRASP) on social cognition and social functioning outcomes.

To summarise the detailed analysis in Appendix A, the RCTs in the field of psychotic disorders seem to be homogeneous to an extent. The RCTs assessed were all pre-registered. The outcome variables of the studies were psychosocial and psychological, with one medication-oriented variable in Article 2 and one physiological variable in Article 8. In all studies participants were recruited from a clinical population with either an ICD-10 (World Health Organization, 1993), DSM-5 (American Psychiatric Association, 2013), DSM-IV (American Psychiatric Association, 1994) or DSM-IV-TR (American Psychiatric

Association, 2000) diagnosis for SSD. However in Articles 1, 2, 3 and 5 nothing is mentioned about establishment of a diagnosis with structured and validated assessments. Most of the eligibility criteria overlap. In all studies participants needed to have an established DSM or ICD diagnosis of a psychotic disorder. In almost all studies the participants had to be between 18 and 65 years old, except for Article 5, 6 and 7. In these studies the maximum age is differing, see Appendix A. The most common exclusion criteria seen in the studies were a neurological disorder, a learning disability, actively abusing substances or a diagnosis of substance abuse disorder. Some of the articles also had other eligibility criteria that did not overlap with other RCTs. For an extensive overview, please see Appendix A.

The studies all had two groups, one control group, with treatment as usual (TAU) as the most commonly used comparison, and one experimental group in which an experimental intervention was to be assessed, with rater-blinded blinding (Articles 1 to 8). For TAU participants were offered an array of different group-therapies, MCT and/or individual CBT. Furthermore, patients are offered occupational therapy, physiotherapy, and physical exercises. In Article 2 the control group consisted of 10 weekly group meetings to share experiences instead of TAU. The interventions tested in Article 1 to 8 were all single interventions and either a combination of psychosocial and psychological interventions or a psychological intervention. In Article 1 to 8 the duration of the intervention varied between four weeks and 18 months. The trial designs were two-armed experimental or quasi-experimental (Article 3), with parallel groups. All studies included at least either a baseline measure and a post-intervention measure. The time between baseline and post-intervention measurements varied in Article 1 to 8, but follow-up measurements were mostly conducted between three and six months after the end of the intervention. Only Article 1, Article 5 and Article 8 did not include a follow-up measurement.

The outcome variables varied across Article 1 to 8, such as mindfulness, psychiatric

symptoms and mentalizing ability. See Appendix A for a full overview of outcome variables. In all studies the outcomes were in line with the aim of the study. A difference across studies was the sample size and gender composition of the groups. Article 1 had a larger sample size, 154 participants to be exact, compared to the other articles (Article 2 to 8), which had 50 or less participants. The gender composition of the groups differed across studies. The instruments and duration varied between studies, given that all eight studies had different outcome variables and interventions used to assess the efficacy of this intervention. Both Article 1 and Article 3 do not report on using the CONSORT guidelines for RCTs. In Article 4, Article 5, Article 6, Article 7 and Article 8 the use of CONSORT is not mentioned, only the flow chart is given in the article. Article 2 does mention the use of CONSORT guidelines for RCTs to assess the quality of the RCT, see Appendix A.

Statistical analyses of RCTs

In most of the articles some form of ANOVA, such as the two-way repeated measures ANOVA and ANCOVA, was used to assess between-group effects. Article 4 used t-tests to assess between-group effects. Article 5 used linear regression to assess between-group effects. Article 6 and Article 8 used a multilevel analysis (MLA) with a linear mixed-effect model to assess between-group effects. Article 7 also included mediation analyses, moderation analyses and MLA with a mixed-effect regression model to assess possible mediation and moderation and to capture differences over time between groups.

Descriptive statistics were used by most studies, just as Chi-square tests and t-tests to assess baseline differences in respectively categorical and continuous demographic data or parametric data. Assumptions for regression, ANOVA or MLA were neither discussed nor checked in the studies. Removal of participants from the analysis was mostly due to drop out or missing data. No outliers were reported in the studies. Missing data was either imputed with endpoint analysis (Article 1), pro-rating (Article 5), maximum likelihood approach

(Article 6 and 8), multiple imputation (Article 7) or participants were deleted before continuing the analysis, see Appendix A.

Results and conclusions

All studies reported a positive trial, see Appendix A. The most commonly reported effect size is the standardised effect size, such as partial eta squared (η2), which is used in Article 2, 3 and 7 and Cohen's *d* (Article 2, 4 and 6). Other standardised effect sizes are the Least Squares Means (Article 6), Cohen's *F* (Article 1) and SES (Article 8). Article 5 used an unstandardised effect size; the adjusted mean difference. Only Article, 1, 2, 3 and 7 interpreted effect sizes. Article 1 reported small to large effect sizes, whereas Article 3 and 7 reported medium to large effect sizes. Article 2 reported large effect sizes. For future research, Article 1 suggests the addition of a placebo group and a 3-month or 6-month follow-up measurement in future research. Article 2 suggests that future research in this area should focus on a fully powered trial to further assess the efficacy of MBGT. Article 3 suggests the use of bigger sample sizes in future research to further assess the effectiveness of MCT. Article 4 has no suggestions for follow-up research. Article 5 suggested the inclusion of a longer-term follow-up, greater emphasis on adverse life events beforehand, extending the therapy window from 6 to 9 months and the re-assessment of instruments used to capture voice-related targets and a more sensitive scale for assessing voice severity. Article 6 suggests that future research focuses on the mechanisms of I-CAT that improve negative symptoms and psychosocial functioning, includes mindfulness and positive psychology, test a group format or digital intervention and that larger scale RCTs are to be conducted. Article 7 suggests lengthening the duration of the treatment in future research. Article 8 suggests the use of a fully powered trial, the focus on developing renewed interventions, focus on changing feelings in social situations, focus on good quality evidence and the evaluation of GRASP against other interventions.

Quality review

The quality differs between articles. Overall, the operationalization follows logically from the research question and rationale of the intervention. Only Article 2, 4, 6 and 7 had clear hypotheses on the effect of the intervention. In Article 1, 2 and 3 justification for the sample size was given by using power calculations or by following the sample size recommendations of other relevant articles. Articles 4, 5, 6, 7 and 8 did not mention the power and had no justification for sample size. The groups included in the RCTs are heterogeneous and no significant between-group differences on demographic data were found between groups, see Appendix A. However, there were some differences between the control group and intervention group in Article 2; the intervention group took higher doses of antipsychotic medication at both baseline and post-intervention measurements. Overall the conditions for both intervention groups and control groups were defined and sessions were summarised for reproducibility. Whether the reliability statistics were given varied across studies. For most of the instruments used in Articles 1 and 3 reliability statistics were given. However, there were some instruments that did not have at least an acceptable reliability or validity, such as the HoNOS (Article 1) and the ToM-1 and ToM-2 (Article 3). Article 2 gave reliability statistics for only two instruments. Articles 4, 5, 6,7 and 8 did not give reliability statistics at all. The amount of measurement points were stated. All eight articles reported results and (un)standardised effect sizes to quantify the impact of the intervention on the outcome variables. The clinical implication was explained in all articles. Also, suggestions for follow-up research were given.

What stood out was the either poor or missing justification of methodological choices in all eight articles. The majority of the chosen eligibility criteria was not justified, which means no argumentation was given, see Appendix A. Only Article 2 justified the exclusion of patients with a score of six or more on the positive scale of the PANSS. Also, the choice for a

specific number of sessions or duration of the therapy was not explained, just as the choice for a specific trial design. Furthermore, none of the articles explained their reasoning for choosing the statistical analysis of choice. However, this judgement is subjective. In addition to this, no assumption checks were reported when conducting an ANOVA, regression or MLA. The reason for participant removal was justified in all eight studies and the method of dealing with missing data also. However, it is unclear how Article 2 dealt with missing data.

Discussion

Study results in light of current research

In this review the (a) methodological and (b) statistical choices as well as (c) the results and interpretation of the results and (d) the quality of the different methodological and statistical aspects of RCTs on psychological interventions in the field of psychotic disorders were assessed to find out what randomised controlled trials of psychological interventions for psychotic disorders look like in terms of methodology and analysis. Differences in methodology and data-analysis were expected, for example differences in sample size, trial design, dealing with assumptions and violations and the statistical analysis used to examine the data. Because all RCTs were using a different methodology and statistical analysis, differences in quality can be observed between the RCTs included.

To summarise, the RCTs in the field of Psychotic Disorders are homogeneous to an extent. RCT's look alike in terms of participant inclusion, trial designs, group design and types of outcome variables. The resemblance on inclusion of participants and group design is to be expected, because most of the RCTs are interested in the efficacy and feasibility of the intervention in a clinical population (Ivarsson & Andersen, 2016). Additionally, RCTs have a rigid methodology and set rules such as the inclusion of a control group, which can explain why RCTs are homogeneous to an extent (Ivarsson & Andersen, 2016). However, the similarities in trial design between the studies was not expected. Trial design in RCTs is

dependent on the context of the study (Correll et al., 2022). Given that there are a multitude of ways to shape the trials, it was expected that the RCTs included in the final review had differences in trial design. If more articles were to be included in the final review, the differences in trial design might have been bigger.

Even though RCTs are somewhat homogeneous, there still is variation in methodology in for example sample size. Some aspects of RCTs, such as the design, cannot be standardised because this is a contextual factor depending on the aim of the study, which explains the variation (Correll et al., 2022). This variation was expected, but seems to be contradictory to the main selling point of RCTs, which is standardisation of research. Even though RCTs strive to standardise, there are no set rules for the sample size (Ivarsson & Andersen, 2016). If unaware of these differences, the importance of this factor on the outcome might be unwittingly ignored and as a result of that results might get overinterpreted. In a smaller sample size it is easier to find an effect and studies with small sample sizes have less power (Franco et al., 2014). This can also be reflected back to the file drawer effect, which tells that articles with positive results are more likely to be published but also more likely to contain false positives because of for example small sample sizes and low power (Lemaire et al., 2022; see also Franco et al., 2014). Furthermore the difference in intervention is due to the variety of interventions available to improve symptoms in individuals with SSD and part of the goal of testing different interventions with RCTs (Correll et al., 2022).

RCTs also look alike in terms of statistical analysis. Most studies reported using ANOVA. The use of ANOVA in most of the RCTs included was not expected, but can be explained. ANOVA has many applications and is widely used in clinical trials to assess the differences between two groups (Singh et al., 2013; see also Vickers, 2005). Given the small scope of this review and the widespread use of ANOVA in clinical trials, it is likely that most of the articles assessed have used ANOVA for statistical analysis of the data. ANOVA can be

used to compare groups, but maybe in light of EBP and the importance of personal characteristics it might be more fitting to compare individual results to see what is working for the individual patient instead of what is working for a group (Bakker, 2019).

What was striking about the study outcomes is that all studies reported a positive trial. This can be explained by the file drawer effect. Trials that report negative or neutral results are less likely to be published, whereas positive results, especially from small trials, are overrepresented (Lemaire et al., 2022). Almost all RCTs that were reviewed had a small sample size, so it is likely that the effect that the authors have found does not exist in real life (Franco et al., 2014). Based on the aim and results of the RCTs reviewed, one can say that the interventions tested are effective in reducing symptoms of mental disorders, which can be plausible because the mere purpose of RCTs is to find out what is working for these disorders. However, this can also be alarming. In clinical practice symptom reduction or removal and therefore 'curing' the individual may not be the aim of the treatment (Bakker, 2019; see also Bakker, 2022). The aim of treatment must be targeting psychological processes (Bakker, 2019). RCTs might lose their purpose of finding out what works in clinical practice if one strives for positive results and standardised steps get followed blindly (Bakker, 2019; Carey et al., 2017).

The quality of the RCTs differed, but what stood out in all articles was the poor or missing justification of methodological and statistical choices. Most of the RCTs that were analysed also had poor reporting. The poor reporting in RCTs is a known issue and can lead to difficulties understanding the study and difficulties in reproducing the study (Vinkers et al., 2021). After checking the assumptions for a statistical analysis, data transformation is sometimes necessary in case of violation of assumptions and now it is not clear whether important assumptions were violated and if changes were made to the data (Lee & Tse, 2017). If one were to be fully reproducing the study, it would be unclear how the data needs to be

analysed (Lee & Tse, 2017). Then, take for example statistical analysis. For readers or clinicians with a background in statistics, the choice for ANOVA may be obvious and may not be needing additional explanation. However, not every clinician has an extensive knowledge of statistics, which makes it difficult to understand the choices that have been made in the RCTs regarding the statistical analysis.

As a master student and aspiring clinician, and thereby part of the target audience of RCTs, I found it quite difficult to follow why some authors chose to do an ANOVA and some authors chose to do a MLA. I felt strongly that this article was written for an audience who has extensive knowledge on the subject and thus choices are deemed obvious. Personally, the choice for either of the analyses is not as obvious given the multitude of options available for analysing group data. The absence of justification left me wondering whether I should have known why the authors chose to do that analysis. It made me doubt my own knowledge. Also, this might be repelling and make people read less trials.

The poor or non-existent justification of the choices that have been made in the RCT may be due to the fact that the authors have limited journal space and words when writing an article that is to be published (McErlean et al., 2023). Most journals have a word limit on articles, which makes it difficult to describe everything into detail and elaborate on their choices (McErlean et al., 2023). Because of the word limit in journals, the information given in the articles is sometimes not concise enough. To tackle this problem, some authors do include detailed information and tables in corresponding online repositories, but it is unlikely that clinicians will look that up.

Moreover, there is variation in the reporting of RCTs, without proper reason or explanation. This is also a known issue (Lloyd-Evans et al., 2014). Even though some of the RCTs have claimed to have used CONSORT for their reporting, the reports of these two articles still varied. Furthermore some articles had methodological flaws, such as baseline

non-equivalence (Article 2), meaning that the groups differ on baseline characteristics after randomisation, which is a common problem in RCTs (Steeger et al., 2021). The influence of confounding variables cannot be ruled out when groups differ on baseline characteristics (Steeger et al., 2021).

The poor or non-existent justification of choices and reporting of RCTs, combined with methodological flaws and the variation in the reporting of RCTs shows the importance of clinicians critically looking at RCTs. We only had group-comparing RCTs. Be aware that that is not always transferable to the individual patient (Bakker 2019). Group comparison shows small effects for many people. It is advised to also consider other evidence, such as case studies and longitudinal research to show individual results and results over time, which RCTs generally do not show. The highest-quality status does not imply quality and thus is not the automatic best fit for the individual patients clinicians see in daily clinical practice (Carey et al., 2017). It is therefore recommended to stay critical and combine information on the efficacy of treatments from multiple primary and secondary sources of evidence to create the best fitting treatment for that patient (Carey et al., 2017). It is advised to stay aware of the possible file drawer effect in research when reading the articles, especially with studies that report positive results but have small sample sizes and less power. Another recommendation would be for clinicians to stay in touch with current research to know about the value of different studies, but also give input to make research practice-relevant. This can be done by participating in participatory action research at local institutions or universities (Cornish et al., 2023) and by following an established training for empirically supported treatments (Teachman et al., 2012)

Limitations

One of the limitations of this study is that the exclusion of psychiatry in the search string may have caused a loss of eligible articles. Even though SSD is a psychiatric mental

disorder, patients will still receive some sort of psychological treatment next to their medication. The studies that have been conducted in the field of psychiatry are thus also relevant for this review. This methodological flaw can affect the reliability of the review, because it can lead to biassed outcomes (Steeger, 2021).

Another limitation was the subjective quality review that was used to judge the quality of the RCTs included. All relevant subjects included in the quality review have been drawn up in the data-extraction table and the quality review and its focus was reviewed several times by both the daily supervisor and two other scientists with extensive knowledge of the subject. Based on these reviews, a consensus was reached about the contents of the quality review. Nevertheless the quality review still remains subjective, given that all judgement is based on what we consider important to maintain a high-quality status, which may have been biassed or can differ from what other clinicians and scientists think is important to determine the quality of an article. This can threaten the validity and reliability of the review, given the possibility that the quality review may not measure what it intends to measure and its risk for biassed outcomes (Steeger, 2021).

The review also had a small scope. Unfortunately this meant that the quality was judged based on whether or not choices were justified, but this does not warrant an in-depth critical reflection of the quality of the RCTs. Additionally, the implications of the different methodological and statistical choices, just as the implications of the quality review that was conducted, could be described and reflected on more in depth. In this review we have identified a problem, but have not yet elaborated on why it is a problem and what the consequences of this problem are.

The small scope also substantially limited the amount of articles that were analysed and included in the final review. It may be that the articles found showed little variation between them, while the RCTs are actually quite diverse. This means there is a possibility

that the results and conclusion of this study are not generalizable to all RCTs in the field of psychotic disorders and its validity is compromised. Additionally, only two databases were used to search for eligible articles, which means there is a possibility that not every eligible article was included in the screening. However, if more articles were to be included in the screening, then it would not have been feasible.

Suggestions for future research

It is recommended that for future research into this topic the scope of the article is broadened to include more databases, more articles and a critical reflection on the implications of the different methodological and statistical choices, as well as a critical in-depth reflection on the quality of the different RCTs. It is also recommended that the part of the search string that involves the exclusion of psychiatry is removed to enable the inclusion of more eligible papers.

Other suggestions for follow-up research are reviewing the eligibility criteria and revising them. Some articles had included participants that were between the ages of 16 and 71, but the participants' age mostly fell between 18 and 55 years old. In this review only studies who included participants between the ages of 18 and 65 were included, but that criteria may have been too harsh considering some studies were excluded based on this criteria, even though the age of the participants fell mostly between 18 and 55 years old.

Conclusion

In this review, I wanted to create an overview of how randomised controlled trials of psychological interventions for psychotic disorders look like in terms of methodology and analysis. To conclude, RCTs are homogeneous to an extent; there is similarity in participant inclusion, trial designs and group design. The RCTs were experimental with parallel groups. RCTs had two groups: one experimental and one control. Most methodological components varied across RCTs. The statistical analyses that were used are similar across trials, which are

ANOVA, regression, t-tests and MLA. All studies reported a positive trial. The quality of the RCTs varied across studies and most RCTs had poor or missing justification of methodological and statistical choices and poor reporting.

Suggestions for further research are broadening the scope of the review to include more databases, more articles and a critical reflection on the implications of the different methodological and statistical choices, as well as a critical in-depth reflection on the quality of the different RCTs. Another suggestion is the reviewing and revising of the eligibility criteria. For daily clinical practice it is recommended to combine information on the efficacy of treatments from multiple primary and secondary sources of evidence to get an integrated framework that contains recommendations for treatment from multiple sources, which in turn can be integrated with the patient's personal characteristics to create the best fitting treatment for that patient. Lastly, remember to stay critical and reflect on what you are reading.

References

Akobeng, A. K. (2005). Understanding randomised controlled trials. *Archives of disease in childhood*, *90*(8), 840-844. https://doi.org/10.1136/adc.2004.058222

Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gotzsche, P. C., Lang, T., & CONSORT Group. (2001). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine, 134*(8), 663–694. https://doi.org/10.7326/0003-4819-134-8-200104170-00012

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.).

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.).

American Psychiatric Association, Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*(4), 271–285. https://doi.org/10.1037/0003-066X.61.4.27

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Bakker, G. M. (2019). A new conception and subsequent taxonomy of clinical psychological problems. *BMC psychology*, *7*, 1-33. https://doi.org/10.1186/s40359-019-0318-8

Bakker, G. M. (2022). Psychotherapy outcome research: Implications of a new clinical taxonomy. *Clinical Psychology & Psychotherapy*, *29*(1), 178-199. https://doi.org/10.1002/cpp.2638

Becker, C. B., Stice, E., Shaw, H., & Woda, S. (2009). Use of empirically supported interventions for psychopathology: can the participatory approach move us beyond the research-to-practice gap?. *Behaviour research and therapy*, *47*(4), 265–274. https://doi.org/10.1016/j.brat.2009.02.007

Brendel, R. W., & Stern, T. A. (2005). Psychotic symptoms in the elderly. *Primary Care Companion to the Journal of Clinical Psychiatry*, *7*(5), 238. https://doi.org/10.4088/pcc.v07n0506

Carey, T. A., & Stiles, W. B. (2016). Some problems with randomised controlled trials and some viable alternatives. *Clinical Psychology & Psychotherapy, 2*3(1), 87-95. https://doi.org/10.1002/cpp.1942

Carey, T. A., Tai, S. J., Mansell, W., Huddy, V., Griffiths, R., & Marken, R. S. (2017). Improving professional psychological practice through an increased repertoire of research methodologies: Illustrated by the development of MOL. *Professional Psychology: Research and Practice*, *48*(3), 175. https://doi.org/10.1037/pro0000132

Cartwright, N. (2011). A philosopher's view of the long road from RCTs to effectiveness. *The Lancet*, *377*(9775), 1400-1401. https://doi.org/10.1016/s0140-6736(11)60563-1

Cella, M., Sedgwick, O., Lawrence, M., Grant, N., Tsapekos, D., Harrison, L., & Wykes, T. (2023). Evaluating the mechanisms of social cognition intervention in schizophrenia: A proof-of-concept trial. *Psychiatry research*, *319*, 114963. https://doi.org/10.1016/j.psychres.2022.114963

Cornish, F., Breton, N., Moreno-Tabarez, U., Delgado, J., Rua, M., de-Graft Aikins, A., & Hodgetts, D. (2023). Participatory action research. *Nature Reviews Methods Primers*, *3*(1), 34.

Correll, C. U., Kishimoto, T., & Kane, J. M. (2022). Randomized controlled trials in schizophrenia: opportunities, limitations, and trial design alternatives. *Dialogues in clinical neuroscience*. https://doi.org/10.31887/DCNS.2011.13.2/ccorrell

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social science & medicine*, *210*, 2-21. https://doi.org/10.1016/j.socscimed.2017.12.005

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502-1505. https://doi.org/10.1126/science.1255484

GBD 2019 Mental Disorders Collaborators (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *The lancet. Psychiatry*, *9*(2), 137–150. https://doi.org/10.1016/S2215-0366(21)00395-3

Guidi, J., Brakemeier, E. L., Bockting, C. L. H., Cosci, F., Cuijpers, P., Jarrett, R. B., Linden, M., Marks, I., Peretti, C. S., Rafanelli, C., Rief, W., Schneider, S., Schnyder, U., Sensky, T., Tomba, E., Vazquez, C., Vieta, E., Zipfel, S., Wright, J. H., & Fava, G. A. (2018). Methodological Recommendations for Trials of Psychological Interventions. *Psychotherapy and psychosomatics*, *87*(5), 276–284. https://doi.org/10.1159/000490574

Halverson, T. F., Meyer-Kalos, P. S., Perkins, D. O., Gaylord, S. A., Palsson, O. S., Nye, L., Algoe, S. B., Grewen, K., & Penn, D. L. (2021). Enhancing stress reactivity and wellbeing in early schizophrenia: A randomized controlled trial of Integrated Coping Awareness Therapy (I-CAT). *Schizophrenia research*, *235*, 91–101. https://doi.org/10.1016/j.schres.2021.07.022

ten Have, M., Tuithof, M., van Dorsselaer, S., Schouten, F., Luik, A. I., & de Graaf, R. (2023). Prevalence and trends of common mental disorders from 2007-2009 to 2019-2022: results from the Netherlands Mental Health Survey and Incidence Studies (NEMESIS), including comparison of prevalence rates before vs. during the COVID-19 pandemic. *World psychiatry: official journal of the World Psychiatric Association (WPA)*, *22*(2), 275–285. https://doi.org/10.1002/wps.21087

Hayes, S. C., & Hofmann, S. G. (2017). The third wave of cognitive behavioral therapy and the rise of process-based care. *World psychiatry: official journal of the World Psychiatric Association (WPA)*, *16*(3), 245–246. https://doi.org/10.1002/wps.20442

Hiebert, R., & Nordin, M. (2006). Methodological aspects of outcomes research. *European spine journal: official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society,* 15(Suppl 1), S4–S16. https://doi.org/10.1007/s00586-005-1057-5

Hunsley, J. (2015). Translating evidence-based assessment principles and components into clinical practice settings. *Cognitive and Behavioral Practice*, *22*(1), 101-109. https://doi.org/10.1016/j.cbpra.2014.10.001

Ivarsson, A., & Andersen, M. B. (2016). What counts as "evidence" in evidence-based practice? Searching for some fire behind all the smoke. *Journal of Sport Psychology in Action*, *7*(1), 11-22. https://doi.org/10.1080/21520704.2015.1123206

Karanicolas, P. J., Farrokhyar, F., & Bhandari, M. (2010). Practical tips for surgical research: blinding: who, what, when, why, how?. *Canadian journal of surgery, 53*(5), 345–348.

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Üstün, T. B. (2007). Age of onset of mental disorders: a review of recent literature. *Current opinion in psychiatry*, *20*(4), 359-364. https://doi.org/10.1097/YCO.0b013e32816ebc8c

Lee, P. H., & Tse, A. C. Y. (2017). Diagnosis checking of statistical analysis in RCTs indexed in PubMed. *European journal of clinical investigation*, *47*(11), 847–852. https://doi.org/10.1111/eci.12833

Lemaire, B. S., Lenoble, R., Zanon, M., Jacquel, T., & Vallortigara, G (2022). The file drawer effect–a long-lasting issue in the sciences. *In&Vertebrates*.

Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical psychology review*, *33*(7), 883-900. https://doi.org/10.1016/j.cpr.2012.09.008

Lloyd-Evans, B., Mayo-Wilson, E., Harrison, B., Istead, H., Brown, E., Pilling, S., Johnson, S., & Kendall, T. (2014). A systematic review and meta-analysis of randomised controlled trials of peer support for people with severe mental illness. *BMC psychiatry*, *14*, 39. https://doi.org/10.1186/1471-244X-14-39

McErlean, M., Samways, J., Godolphin, P. J., & Chen, Y. (2023). The reporting standards of randomised controlled trials in leading medical journals between 2019 and 2020: a systematic review. *Irish journal of medical science*, *192*(1), 73–80. https://doi.org/10.1007/s11845-022-02955-6

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan — a web and mobile app for systematic reviews. *Systematic Reviews, 5*, 210. https://di.org/10.1186/s13643-016-0384-4

Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical psychology review*, *71*, 90–100. https://doi.org/10.1016/j.cpr.2019.01.002

Reddy, Y. C. J., Sudhir, P. M., Manjula, M., Arumugham, S. S., & Narayanaswamy, J. C. (2020). Clinical Practice Guidelines for Cognitive-Behavioral Therapies in Anxiety Disorders and Obsessive-Compulsive and Related Disorders. *Indian journal of psychiatry*, *62*(Suppl 2), S230–S250. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_773_19

Shafran, R., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A., ... & Wilson, G. T. (2009). Mind the gap: Improving the dissemination of CBT. *Behaviour research and therapy*, *47*(11), 902-909. https://doi.org/10.1016/j.brat.2009.07.003

Singh, V., Rana, R. K., & Singhal, R. (2013). Analysis of repeated measurement data in the clinical trials. *Journal of Ayurveda and integrative medicine*, *4*(2), 77–81. https://doi.org/10.4103/0975-9476.113872

Steeger, C. M., Buckley, P. R., Pampel, F. C., Gust, C. J., & Hill, K. G. (2021). Common Methodological Problems in Randomized Controlled Trials of Preventive Interventions. *Prevention science: the official journal of the Society for Prevention Research*, *22*(8), 1159–1172. https://doi.org/10.1007/s11121-021-01263-2

Stronks, K., Wieringa, N. F., & Hardon, A. (2013). Confronting diversity in the production of clinical evidence goes beyond merely including under-represented groups in clinical trials. *Trials*, *14*, 177. https://doi.org/10.1186/1745-6215-14-177

Tackett, J. L., Brandes, C. M., King, K. M., & Markon, K. E. (2019). Psychology's replication crisis and clinical psychological science. *Annual review of clinical psychology, 15*, 579-604. https://doi.org/10.1146/annurev-clinpsy-050718-095710

Teachman, B. A., Drabick, D. A., Hershenberg, R., Vivian, D., Wolfe, B. E., & Goldfried, M. R. (2012). Bridging the gap between clinical research and clinical practice: introduction to the special section. *Psychotherapy (Chicago, Ill.)*, *49*(2), 97–100. https://doi.org/10.1037/a0027346

Vickers A. J. (2005). Analysis of variance is easily misapplied in the analysis of randomized trials: a critique and discussion of alternative statistical approaches. *Psychosomatic medicine*, *67*(4), 652–655. https://doi.org/10.1097/01.psy.0000172624.52957.a8

Vinkers, C. H., Lamberink, H. J., Tijdink, J. K., Heus, P., Bouter, L., Glasziou, P., Moher, D., Damen, J. A., Hooft, L., & Otte, W. M. (2021). The methodological quality of

176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. *PLoS biology*, *19*(4), e3001162. https://doi.org/10.1371/journal.pbio.3001162

Weijers, J., Ten Kate, C., Viechtbauer, W., Rampaart, L. J. A., Eurelings, E. H. M., & Selten, J. P. (2021). Mentalization-based treatment for psychotic disorder: a rater-blinded, multi-center, randomized controlled trial. *Psychological medicine*, *51*(16), 2846–2855. https://doi.org/10.1017/S0033291720001506

World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders*. World Health Organization.

World Health Organization. (2022). *ICD-11: International classification of diseases* (11th revision). https://icd.who.int/

Zabor, E. C., Kaizer, A. M., & Hobbs, B. P. (2020). Randomized controlled trials. *Chest*, *158*(1), S79-S87. https://doi.org/10.1016/j.chest.2020.03.013

Zisman-Ilani, Y., Chmielowska, M., Dixon, L. B., & Ramon, S. (2021). NICE shared decision making guidelines and mental health: challenges for research, practice and implementation. *British Journal of Psychiatry open*, *7*(5), e154. https://doi.org/10.1192/bjo.2021.987

Appendix A

Data extraction table and Tables for the data extraction of all articles

Table 1

Data extraction table

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
Methodology			
Pre-registration of the article	Pre-registration No pre-registration	Was the article pre-registered?	
Outcome variable	Psychological Psychosocial Cognitive Other	Does the operationalization follow from the research question and the rationale of the intervention?	
Sample size	N	Was there any justification of sample size?	
Participants	Clinical group (diagnosed) Non-clinical group Gender ● Male ● Female ● Mixed group	Homogenous or heterogenous Justification of inclusion/exclusion of participants	
Groups	Number of experimental groups Control group ● Waiting list ● TAU (treatment as usual) ● other	Clearly defined conditions	
Instruments	Self-report measurement Test Clinical interview Semi-structured interview	Reliability statistics given	
Inclusions	Age Comorbidity Symptoms DSM-diagnosis Gender	Clearly defined argumentation for including or excluding	

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
Intervention	Single/combination Psychological/Psychosocial	Clearly defined type of intervention	
Duration of therapy	Number of sessions Length	It is clear how long the treatment lasted and why	
Trial design	Parallel group Crossover Cluster Factorial Single-group Non-randomized comparative Other	Are the decisions and choices justified?	
Measuring points	Number of measurement points Baseline Pre-treatment Post-treatment Follow-up	Stated the number of measurement points and in what timeframe the measurements took place	
Statistical analysis			
Statistical analysis	T-test Descriptive ANOVA ANCOVA Simple regression multiple regression Repeated measures ANOVA Mediation variables included Moderation variable included Multilevel analysis	Is the model justified? Changes in planning are justified	
Removed participants	Due to drop-out Did not meet criteria (after conducting research) Due to missing data	The reason of removal is justified	

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
	Due to spurious pattern		
Assumptions (does not concern all statistical analyses)	Homogeneity of variance Normal distribution of residuals Linearity Independent observations	Are the assumptions stated? Assumptions are met and if violated the issue is resolved, or justified why no changes were conducted	
Outliers	Outlier removal rule	Outliers are removed if necessary and reasonable argumentation has to be given	
Missing data	Data imputation Participant deletion	Clearly described if and why data is imputed or a participant is deleted	
Results & Conclusion			
Result	Positive trial Negative trial Neutral trial Outcome in line with aim study: yes/no	Reported in conclusion and abstract	
Effect size	Unstandardized effect size Cohen’s d Hedges g Eta squared Partial eta squared Omega squared Cohen’s W Cramer's V Pearson correlation Cohen’s F F-squared R-squared	Effect size, if given, is interpreted	

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
Conclusion	Clinical relevance Impact on daily practice	Is anything said about relevance at all?	

Table 2

Data extraction table Article 1, Positive Psychotherapy Psychosis (Hon-wai Chu et al., 2021)

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
Methodology			
Preregistration of the article	The article was preregistered	-	
Outcome variable	Psychological: • Psychiatric symptom severity • Depressive symptoms Psychosocial: • Mental well-being • Hope • Self-efficacy • Quality of life • Social disability	Yes, the operationalization follows from the research question and rationale of the intervention	
Sample size	*n* = 154	Justification was given based on power and alpha	
Participants	Clinical group • ICD-10 diagnosis F20 - F29 • Mixed group, mostly female	Groups are heterogenous and intervention and control group do not differ much	

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
Groups	Two groups • One control group with treatment as usual (TAU) • One experimental group with positive psychotherapy for psychosis (PPP)	Conditions for each group are clearly defined	

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
Instruments	Self-report: ● The *Chinese Short Warwick-Edinburgh Mental Well-being Scale* (CSWEMWS) ● The *Chinese Hope Scale* (CHS) ● The *Chinese General Self-efficacy Scale* (CGSS) ● The *Short Form-12* (SF-12) Clinical interview: ● The *Health of the Nation Outcome Scale* (HoNOS) ● The *Brief Psychiatric Rating Scale* (BPRS) ● The *Cantonese version of the Calgary Depression Rating Scale* (CDSS-C)	The CSWEMWS had a good internal consistency (α = .89), just as the CGSS (α = .92) and the CDSS-C (α = .80). The CHS had an acceptable internal consistency (α = .70). The HoNOS had a questionable internal consistency, which varied from α = .59 to α =.76. The BPRS had an acceptable to good internal consistency, which varied from α = .65 to α = .88. The test-retest reliability was good to excellent for the CGSS, BPRS and CDSS-C, with respectively *r* = .85, *r* = .87 to .97 and *r* = .86. However, the test-retest reliability of the HoNOS was	

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
		poor to acceptable (*r* = .31 to .65)	

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
Eligibility	Inclusion: • Aged 18 to 65 years old • Comorbidity is not stated in eligibility • ICD-10 diagnosis F20 to F29 • Using secondary health services • Need to read Chinese and speak Cantonese • Written consent Exclusion: • Cognitive impairment • Active substance abuse • No consent	No argumentation given for chosen inclusion and exclusion criteria	
Intervention	Single Combination of a psychological and psychosocial intervention.	The intervention is clearly defined, and every session is summarised	
Duration of therapy	13 sessions over the course of 7 consecutive weeks	Nothing is stated about the choice for or the efficacy of the specific number of sessions	
Trial design	Parallel group, two-armed design	No justification of trial design and	

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
		choices that have been made.	
Measuring points	Two measurement points: • Baseline • Post-treatment	Stated the number of measurement points and in what timeframe the measurements took place	
Statistical analysis			
Statistical analysis	Chi-square: • Categorical demographic data T-test: • Continuous demographic data • Baseline assessments on outcome measure Two-way repeated measures ANOVA: • Within patient • Between groups • Over time	No justification for the chosen method of analysis and no choices or changes in plan are specified	
Removed participants	Due to drop-out Due to missing data	Reason of removal is justified	

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
Assumptions	The authors did not check assumptions.	No assumptions stated and no information available to judge whether assumptions are met or violated	
Outliers	No outliers	No argumentation needed	
Missing data	Data imputation with endpoint analysis	Reason of imputation is stated: lost to follow-up	
Results & Conclusion			
Result	Positive trial Outcome in line with aim study: yes	Results are reported in the conclusion and abstract	
Effect size	Standardised effect size: ● Cohen's *F* (the effect size was not interpreted)	Statistical significance was interpreted but not effect size	

	Data Choices	Quality Review	
Extracted	**Possible Outcomes**	**Outcome Justification**	**Notes**
Conclusion	Hon-wai Chu et al. (2021): ● Positive psychology-based interventions positively impact well-being but also target other domains well beyond well-being. ● Positive psychology based interventions have a possible protective effect which can delay onset or lower the risk of onset of schizophrenia. ● Positive psychology based interventions can help build resilience. ● This highlights the importance of a new form of positive psychology for SSD Suggestions for future research: ● Measurement at post-intervention intervals such as 3-month or 6-month post-intervention ● Placebo therapeutic group	Relevance to clinical practice and research is stated	

Table 3

Data extraction table for Article 2, Mindfulness Based Group Therapy (Böge et al., 2021)

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Methodology			
Preregistration of the article	The article was pre registered	-	
Outcome variable	Psychosocial: ● Mindfulness ● Social functioning ● Quality of life Psychological: ● Positive and negative symptoms ● Depression ● Anxiety ● Psychological flexibility Other: ● Medication regime at baseline, post-intervention, and follow-up	Yes, the operationalization follows from the research question and rationale of the intervention	
Sample size	$n = 40$	Justified using relevant literature	
Participants	Clinical group ● ICD-10 or DSM-5 diagnosis SDD ● Mixed group	Groups are heterogenous. There were statistical significant differences in dosages of medication; TAU	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
		condition received higher doses of antipsychotic medication	
Groups	Two groups: ● One control group with treatment as usual (TAU) ● One experimental group with Mindfulness Based Group Therapy (MBGT)	Conditions for each group are clearly defined	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Instruments	Self-report: ● The *Comprehensive Inventory of Mindfulness Experiences* (CHIME) ● the *Freiburger Mindfulness Inventory* (FMI) ● The *Depression Anxiety Stress Scale* (DASS) ● The *Cognitive Fusion Questionnaire* (CFQ) ● the *Acceptance and Action Questionnaire* (AAQ-II) ● The *World Health Organization Quality of Life BREF* (WHO-QOL-BREF) Clinical interview: ● The *Positive and Negative Syndrome Scale* (PANSS) ● *Calgary Depression Scale for Schizophrenia* (CDS) ● The *Personal and Social Performance Scale* (PSP)	Not much reliability statistics are given, just their interpretation according to other literature. The only statistics available are test-retest reliability of the PSP (r = 0.79) and the internal consistency of the WHO-QOL-BREF, (a = >.70)	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Eligibility	Inclusion criteria: ● Currently in-patient ● Between 18 and 65 years of age ● DSM-5 and ICD-10 diagnosis SSD ● Informed consent ● Willing and able to follow group therapy Exclusion criteria: ● > 6 on any item on the positive scale of the PANSS ● Suicidality ● Neurological disorders ● Substance abuse	Only the choice for the exclusion criteria concerning positive symptoms has been justified	
Intervention	Single MBGT is a combination of a psychosocial and psychological intervention.	The intervention is clearly defined, and every session is summarised	
Duration of therapy	4 weeks in-patient treatment for treatment as usual (TAU) 4 weeks for MBGT	Nothing is stated about the choice for or the efficacy of the specific number of sessions	
Trial design	Parallel group, two-armed design	No justification of trial design and choices that have been made	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Measuring points	Three measuring points: ● Baseline ● Post-intervention ● Follow-up after 12 weeks	Stated the number of measurement points and in what timeframe the measurements took place	
Statistical analysis			
Statistical analysis	Descriptive statistics: ● Recruitment rate ● Protocol adherence ● Retention rate Chi-square tests and t-tests: ● Between-group differences on demographic data ● The dosage of medication at baseline and post-intervention ANCOVA ● Between-group effects Paired sample t-tests: ● Within-group changes between baseline and post-intervention	No justification for the chosen method of analysis and no choices or changes in plan are specified	
Removed participants	Due to drop out	Removal is justified	
Assumptions	The authors did not check assumptions	No assumptions stated and no information available to judge	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
		whether assumptions are met or violated	
Outliers	No outliers	No justification needed	
Missing data	Data of follow-up measurements of 20 participants is missing; not clear whether the data is imputed or deleted	No information available to judge	
Results & Conclusion			
Result	Positive trial Outcome in line with aim study: yes	Results are reported in the conclusion and abstract	
Effect size	Standardised effect sizes: ● Partial eta squared (medium to large effect sizes) ● Cohen's d (small to large effect sizes)	The effect sizes were stated and interpreted according to guidelines	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Conclusion	Böge et al. (2021): ● Based on the results, evidence is given for the feasibility and acceptability of MBGT to treat SSD ● Promising results highlight the possibility of MBGT for SSD in reducing psychiatric symptoms, such as positive symptoms, negative symptoms and affective symptoms ● MBGT can possibly also improve psychological flexibility, quality of life, and social functioning Suggestions for future research: ● Fully powered trial	Relevance to clinical practice and research is stated	

Table 4

Data extraction table for Article 3, Metacognitive Training (Zonp & Bilgin, 2021)

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Methodology			
Preregistration of the article	The article was pre registered	-	
Outcome variable	Psychosocial: ● Social cognition	Yes, the operationalization follows from the research question and rationale of the intervention	
Sample size	$n = 39$	Justification was given based on power and alpha	
Participants	Clinical group ● DSM-5 diagnosis schizophrenia ● Mixed group, mostly male	Groups are heterogenous, but there were slightly more hospitalizations in the MCT group	
Groups	Two groups: ● One control group with group sessions for sharing information and experience ● One experimental group with Metacognitive Training (MCT)	Conditions for each group are clearly defined	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Instruments	Cognitive tests: ● *Facial Emotion Identification and Discrimination Tests* (FEIT/FEDT) ● *False Belief Tasks* (ToM-1 and ToM-2) Self-report: ● The *Reading the Mind in the Eyes Test* (RMET) ● The *Attributional Style Questionnaire* (ASQ)	The Kuder-Richardson 21 value for the FEIT and FEDT was .60. The KR-21 value for the RMET was .59. The KR-21 values for the ToM-1 and ToM-2 were .42 and .11 respectively. The internal consistency of the ASQ was good (a = .71) The test-retest reliability of the FEIT and FEDT was good to excellent (r = .70 to r = .93)	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Eligibility	Inclusion: ● DSM-5 diagnosis schizophrenia ● Between 18 and 65 years old ● No hearing or vision problem ● Willing to participate ● Currently attending the centre activities Exclusion: ● Comorbid neurological psychiatric diagnosis or substance dependency	No eligibility criteria have been justified	
Intervention	Single Psychological intervention	The intervention is clearly defined, and every module is summarised	
Duration of therapy	10 weeks for control group 10 weeks for MCT	Nothing is stated about the choice for or the efficacy of the specific number of sessions	
Trial design	Quasi-experimental, parallel group, two-armed design	No justification of trial design and choices that have been made	
Measuring points	Three measuring points: ● Pre-intervention one week for start intervention ● Post-intervention	Stated the number of measurement points, but not specifically in what	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	• Follow-up after three months	timeframe the measurements took place	
Statistical analysis			
Statistical analysis	Descriptive statistics: • Means and standard deviation • Demographics The independent t-test: • Parametric data Chi-square test: • Difference in demographics The two-way repeated measures ANOVA: • Change over time • Group and time interactions	No justification for the chosen method of analysis and no choices or changes in plan are specified	
Removed participants	Due to drop out	Removal is justified	
Assumptions	The authors did not check assumptions.	No assumptions stated and no information available to judge whether assumptions are met or violated.	
Outliers	No outliers	No justification needed	
Missing data	Participant deletion	Justified based on the reason for missing data	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Results & Conclusion			
Result	Positive trial Outcome in line with aim study: yes	Results are reported in the conclusion and abstract	
Effect size	Standardised effect size: ● Partial eta squared (medium effect)	The effect sizes were stated and interpreted according to guidelines	
Conclusion	Zonp & Bilgin (2021): ● MCT can be effective in improving the ToM and changing attributional styles in patients with schizophrenia ● MCT is culturally sensitive and can be conducted by mental health nurses Suggestions for future research: ● Bigger sample size	Relevance to clinical practice and research is stated	

Table 5

Data extraction table for Article 4, referral to and an appointment with an Acute Inpatient Unit (AIU) clinical psychologist (Solar et al., 2022)

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Methodology			
Pre-registration of the article	Pre-registration	The article was preregistered	
Outcome variable	Psychological: ● Auditory verbal hallucinations ● Distress Other: ● Seeing a community therapist	The operationalization follows logically from the research question and intervention chosen	
Sample size	$n = 31$	No justification of sample size	
Participants	Clinical group: ● An established diagnosis of Schizophrenia ● Mixed group	Homogenous groups The inclusion and exclusion of participants is justified	
Groups	Two groups: ● One control group with TAU ● One experimental group with a referral to an AIU clinical psychologist	Somewhat defined conditions	
Instruments	Self-report measurement: ● *Engagement in Community Therapy* (ECT) ● *Revised Beliefs about Voices Questionnaire* (BAVQ-R), subscale RE Semi-structured interview:	No reliability statistics given	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	● *Psychotic Symptoms Rating Scales* (PSYRATS), subscales AHS and H-DIS		
Eligibility	Inclusion: ● 18 to 65 years old ● experiencing auditory hallucinations ● Established diagnosis for SSD ● Treating AIU psychiatrist judged referral to an AIU clinical psychologist unnecessary Exclusion: ● Dependent on medical care ● Already referred to an AIU clinical psychologist ● Communication barriers	No argumentation for inclusion or exclusion criteria	
Intervention	Single Psychological	The intervention is clearly defined	
Duration of therapy	Unspecified time and length	It is unclear how long the treatment lasted and why	
Trial design	Parallel group, two-armed design	No justification of decisions and choices	
Measuring points	Five measurement points: ● Baseline ● One week in treatment ● One, three and six months post-treatment	Stated the number of measurement points and in what timeframe the measurements took place	
Statistical analysis			

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Statistical analysis	T-test: ● Comparisons between baseline and the last follow-up Descriptive statistics: ● Baseline demographic differences between groups	The model is not justified Changes in planning are justified	
Removed participants	Due to drop-out Due to missing data	The reason of removal is justified	
Assumptions (does not concern all statistical analyses)	None needed for t-test	-	
Outliers	No outliers	-	
Missing data	Participant deletion	Solar et al. (2022) clearly described why a participant was deleted, so justified	
Results & Conclusion			
Result	Positive trial, but results are weak: some not statistically significant or weak decrease Outcome in line with aim study: yes	The results are reported in conclusion and abstract	
Effect size	Standardised effect size: ● Cohen's d (not interpreted)	Effect size is not interpreted	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Conclusion	Solar et al. (2022, p .456): ● This study encouraged and supported 10 participants to see a community therapist six months after discharge ● For daily clinical practice the referral to a community therapy for other patients with schizophrenia and AVH can be beneficial Suggestions for follow up research: ● No suggestions	Relevance is stated	

Table 6

Data extraction table for Article 5, Talking with Voices trial (TwV) (Longden et al., 2021)

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Methodology			
Pre-registration of the article	Pre-registration	The article is preregistered	
Outcome variable	Psychological: ● Voice-hearing ● General clinical presentation Psychosocial: ● Adversity exposure ● Health economics Other: ● Feasibility ● Acceptability ● Therapeutic relationship ●	The operationalization follows logically from the research question and the rationale of the intervention	
Sample size	*n* = 50	Justification for sample size is given	
Participants	Clinical group • Diagnosis ICD for SSD • Mixed group	Homogenous groups The inclusion or exclusion of participants is justified	
Groups	Two groups: ● One control group with TAU ● One experimental group with Talking with Voices (TwV) + TAU	Clearly defined conditions	
Instruments	Self-report measurement: ● the *Voice and You scale* (VAY)	No reliability statistics given	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	● *Subtypes of Voice Hearing Questionnaire* (VHQ) ● *Revised Beliefs about Voices Questionnaire* (BAVQ-R) ● *Questionnaire About the Process of Recovery* (QPR) ● *revised Dissociative Experiences Scale* (DES-R) ● *Revised Life Stressor Checklist* (LSC-R) ● *EQ-5D* ● *Working Alliance Inventory* (WAI) Clinical interview: ● *Positive and Negative Syndrome Scale* (PANSS)		
Eligibility	Inclusion: ● 18 years or older ● Voice hearing: > 1 year ● scored ≥4 on the AHS of PANSS ● ICD SSD ● No medication changes ● Written consent ● Not having structured therapy ● Using secondary care mental health services	Clearly defined argumentation for including or excluding	Oldest particip ant < 65 years

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	● Willing to talk about the voices ● Voices related to personal experiences to talk about it Exclusion: ● Risk of harm ● Not speaking English ● Diagnosis of substance dependence or ASD ● learning disability ● brain injury or illness resulting in psychotic related symptoms ● scoring >5 on the CD of the PANSS ● being homeless		
Intervention	Single Psychosocial	The intervention is clearly defined	
Duration of therapy	26 sessions over the course of 6 months	It is clear how long the treatment lasted but not why.	
Trial design	Parallel group, two-armed design	Yes, the decisions and choices are justified	
Measuring points	3 measurement points: ● Baseline ● One follow-up ● One follow-up after 26 weeks	Stated the number of measurement points and in what timeframe the measurements took place	
Statistical analysis			
Statistical analysis	Descriptive statistics: ● Baseline demographics	Yes, the model is somewhat justified by explaining their choices for regression	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	Linear regression: ● Between-group adjusted mean difference controlling for baseline scores		
Removed participants	Due to drop-out Due to missing data	The reason for removal is justified	
Assumptions (does not concern all statistical analyses)	No assumptions checked	Assumptions not stated, so nothing can be said about violation of the assumptions	
Outliers	No outliers	-	
Missing data	Data imputation: ● Pro-rating	Clearly described if and why data is imputed or a participant is deleted	
Results & Conclusion			
Result	Positive trial Outcome in line with aim study: yes	Results are reported in conclusion and abstract	
Effect size	Unstandardized effect size: ● Adjusted mean differences (not interpreted)	Effect size is not interpreted	
Conclusion	Longden et al. (2021): ● Too underpowered to obtain definitive evidence	Relevance is stated	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	● Results show us that it is possible for clinicians to engage therapeutically in dialogues with voices ● This also highlights the opportunity that these voices can be conceptualised for an individual and can be associated with the life history of the individual ● There is a possibility that relationships are an important aspect of hearing voices. Relationships and their influence should be considered when supporting clients with distressing voices Suggestions for follow-up research: ● Inclusion of a longer-term follow-up ● Sufficiently resourced trials ● Greater emphasis on adverse life events beforehand ● Extending the therapy window from 6 to 9 months ● Re-assessment of		

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	instruments used to capture voice-related targets ● More sensitive scale for assessing voice severity		

Table 7

Data extraction table for Article 6, Integrated-Coping Awareness Therapy (I-CAT) (Halverson et al., 2021)

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Methodology			
Pre-registration of the article	Pre-registration	Yes, the article was preregistered	
Outcome variable	Psychological: ● Positive emotions ● Symptoms ● Mindfulness ● Well-being ● Stress reactivity Psychosocial: ● Quality of life	The operationalization follows logically from the research question and the rationale of the intervention	
Sample size	*n* = 36	No justification of sample size	
Participants	Clinical group ● Diagnosis according to DSM-IV ● Mixed group	Homogenous groups Justification was given for denying participation and participation in the study	
Groups	Two groups: ● One control group with TAU ● One experimental group with I-CAT	Clearly defined conditions for TAU and I-CAT	
Instruments	Self-report measurement: ● *Differential Emotion Scale* (mDES) ● *First Episode Social Functioning Scale* (FESFS) ● The *Perceived Stress Scale* (PSS)	No reliability statistics are given	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	● The *Five Facet Mindfulness Questionnaire* (FFMQ) ● *Daily Stress Inventory* (DSI) ● The *Self-Compassion Scale Short Form* (SCS) ● *Psychological Well-Being Scale* (PWB) Clinical interview: ● The *Positive and Negative Syndrome Scale* (PANSS) Semi-structured clinical interview: ● *Quality of Life Scale* (QLS)		
Eligibility	Inclusion: ● Meeting DSM-IV criteria for SSD ● Between 18 and 35 years old ● IQ greater than 80 ● No current substance use disorder ● No hospitalizations within the past three months ● No history of meditation in the past year	No justification for eligibility criteria	
Intervention	Single Psychosocial	The intervention is clearly described and important aspects are stated	
Duration of therapy	14 to 24 sessions, flexible administration	No explanation for the chosen length or amount of sessions needed	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Trial design	Parallel group, two-armed design	No justification for choice of trial	
Measuring points	Four measurement points: ● Baseline ● Mid-treatment ● Post-treatment ● Three month follow-up	Stated the number of measurement points and in what timeframe the measurements took place	
Statistical analysis			
Statistical analysis	Descriptive: ● Baseline demographics Multilevel analysis: ● Within- and between-group effects over time	The model and the choices that have been made are not justified	
Removed participants	Due to drop-out Due to missing data	The reason for removal is justified in the flowchart	
Assumptions (does not concern all statistical analyses)	No assumptions were checked	The authors did not check assumptions	
Outliers	No outliers	-	
Missing data	Data imputation: ● Maximum Likelihood	Described the imputation and justified why data is imputed	
Results & Conclusion			
Result	Positive trial Outcome in line with aim study: yes	The results are reported in the conclusion and abstract	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Effect size	Standardised effect size: ● Cohen's d (not interpreted)	Effect size is not interpreted	
Conclusion	Halverson et al. (2021): ● The preliminary results highlight great possibilities for I-CAT and show the feasibility and acceptability of I-CAT for SSD. ● Results show that I-CAT successfully reduced psychiatric symptoms such as negative symptoms ● Results also show that I-CAT successfully increased mindfulness and an individual's purpose in life. ● I-CAT also successfully increased the retaining of work and school related skills Suggestions for future research: ● focus on the mechanisms of I-CAT that improve negative symptoms and psychosocial functioning ● Including mindfulness and positive psychology ● Group format or digital format	Relevance to research and clinical practice is stated	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	● Conducting large-scale RCTs		

Table 8

Data extraction table Mentalization Based treatment for psychotic disorder (MBTp) (Weijers et al., 2021)

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Methodology			
Pre-registration of the article	Pre-registration	The article was preregistered	
Outcome variable	Psychological: ● Positive symptoms ● Negative symptoms ● Depression ● Anxiety ● Lack of insight ● Personality organization and somatization of psychopathology ● (Childhood) Trauma Psychosocial: ● Social functioning ● Mentalizing ability ● Theory of Mind ● Quality of life Other: ● Experience sampling variables ● Adherence to pharmacological treatment ● Duration of illness.	The operationalization follows logically from the research question and the rationale of the intervention	
Sample size	$n = 42$	No justification of sample size	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Participants	Clinical group ● DSM-IV-TR diagnosis, established ● Mixed group, mostly male	Homogenous groups The inclusion/exclusion of participants is justified	
Groups	Two groups: ● One control group with TAU ● One experimental group with Mentalization Based Treatment for psychotic disorder (MBTp)	Clearly defined conditions in both groups	
Instruments	Self-report measurement: ● *Manchester Short Assessment of Quality of Life* (MANSA) ● *Dutch short Form of the MMPI* (DSFM) ● *Medication Adherence Questionnaire* (MAQ) Test: ● *Thematic Apperception Test* (TAT) ● *Hinting Task* Other: ● *Social Cognition and Object Relations System* (SCORS) for scoring the TAT ● Digital diary the *'PsyMate'* ● Theory driven analysis on the DSFM	No reliability statistics given	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	● Counting the number of years since onset of disorder Clinical interview: ● *Social Functioning Scale* (SFS) ● *Positive and Negative Syndrome Scale* (PANSS) NL translation Semi-structured interview: ● *Childhood Experience of Care and Abuse* (CECA)		
Eligibility	Inclusion: ● >6 months but <10 years of treatment for a non-affective psychotic disorder ● between 18 and 55 years of age Exclusion: ● Intellectual disability and/or illiteracy ● No knowledge of Dutch ● Addiction with hospitalisation	No justification for eligibility criteria	
Intervention	Single Psychosocial	The intervention is clearly defined and important aspects highlighted	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Duration of therapy	18 months	Not clear why the specific length or amount of sessions was chosen	
Trial design	Parallel group, two-armed design	No justification for decisions and choices	
Measuring points	Four measurement points: ● Baseline ● after 18 months ● Posttreatment ● follow-up after six months	Stated the number of measurement points and in what timeframe the measurements took place	
Statistical analysis			
Statistical analysis	Descriptive statistics: ● Baseline demographics Repeated measures ANOVA: ● Change over time ANCOVA: ● Differences in groups, adjusted for baseline levels Multilevel analysis: ● Between-group changes over time Mediation analysis: ● The mediating effect of mentalizing ability Moderation analysis: ● The moderating effect of severity of childhood trauma, type of personality	The model is justified, but not clear why choices have been made	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	organization, the degree of somatization, adherence to pharmacological treatment, total number of hours of attended sessions, and duration of illness		
Removed participants	Due to drop-out	The reason of removal is justified	
Assumptions (does not concern all statistical analyses)	No assumptions checked	The authors did not check assumptions.	
Outliers	No outliers	-	
Missing data	Data imputation: ● Multiple imputation	Clearly described that data is imputed and why	
Results & Conclusion			
Result	Positive trial Outcome in line with aim study: yes	The results are reported in the conclusion and abstract	
Effect size	Standardised effect size: ● Partial eta squared (large effect)	Effect size is interpreted according to guidelines	
Conclusion	Weijers et al. (2021): ● There is a possibility that	The relevance for research and clinical practice is not clearly stated	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	MBTp is more successful than TAU in improving psychiatric symptoms, social functioning and mentalizing ability Suggestions for follow-up research: ● Lengthening the duration of the treatment in future research		

Table 9

Data extraction table Group Training for Social skills in Psychosis (GRASP) (Cella et al., 2023)

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Methodology			
Pre-registration of the article	Pre-registration	The article was preregistered	
Outcome variable	Psychological: ● Affect ● Symptoms Psychosocial: ● Social functioning ● theory of mind (ToM) ● Global functioning Cognitive: ● Attribution bias ● Recognizing of emotions Other: ● Physiological arousal	The operationalization follows logically from the research question and the rationale of the intervention	
Sample size	$n = 48$	No justification of sample size	
Participants	Clinical group: ● DSM-5, DSM-IV or ICD-10 diagnosis for SSD ● Mixed group	Homogenous groups Including or excluding participants was justified	
Groups	Two groups: ● One control group with TAU ● One experimental group with Group Training for Social Skills in Psychosis (GRASP) + TAU	Clearly defined conditions for the intervention, but not TAU	
Instruments	Self-report measurement:	No reliability statistics are given	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	● *Work and Social Adjustment Scale* (WSAS) ● *Positive and Negative Affect Scale* (PANAS) ● *Ambiguous Intention and Hostility Questionnaire* (AIHQ) Test: ● *Hinting Task* ● *Facial emotion identification test* (FEIT) Clinical interview: ● *Positive and Negative Syndrome Scale* (PANSS) Other: ● *Empatica E4*		
Eligibility	Inclusion: ● SSD diagnosis according to DSM-IV, DSM-5 or ICD-10 ● Between 18 and 65 years old ● Speaks English ● Social functioning: < 20 hours of social contact per week Exclusion: ● Substance abuse disorder ● Medication changes in last six weeks ● Learning difficulties ● No informed consent	No justification for eligibility criteria	
Intervention	Single Psychosocial	The intervention is clearly defined and important aspects are stated	
Duration of therapy	8 weeks	Not clear why the specific length is chosen	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
Trial design	Parallel group, two-armed design	The decisions and choices made in the trial design are not justified	
Measuring points	Two measurement points ● Baseline ● Post-treatment	Stated the number of measurement points and in what timeframe the measurements took place	
Statistical analysis			
Statistical analysis	Descriptive statistics: ● Baseline demographics Multilevel analysis: ● Linear mixed-effect models for between group changes over time	The model is justified, but the choice for MLA is not	
Removed participants	Due to drop-out	The reason for removal is justified	
Assumptions (does not concern all statistical analyses)	No assumptions checked	The authors did not check assumptions	
Outliers	No outliers	-	
Missing data	Data imputation: ● Maximum Likelihood	Clearly described that the data is imputed but not why	
Results & Conclusion			
Result	Positive trial	The results are reported in conclusion and abstract	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	Outcome in line with aim study: yes		
Effect size	Standardised effect size: ● Mean group difference (Medium and large effects)	Effect size can be interpreted with given guidelines, but is not interpreted by the authors	
Conclusion	Cella et al. (2023): ● The overall conclusion is that improving social functioning in individuals with SSD may be complex and time-consuming. ● The results highlight the feasibility and acceptability of GRASP as an intervention for treating SDD ● However, a group format GRASP is not suitable. The social difficulties make it hard for some individuals to participate ● In daily clinical practice it can be beneficial to offer individual sessions Suggestions for follow-up research: ● the use of a fully powered trial ● Renewed focus on developing interventions ● Changing the feelings in social situations ● Good quality evidence	The relevance for research and clinical practice is stated	

	Data Choices	Quality Review	
Extracted	**Outcomes**	**Outcome Justification**	**Notes**
	● the evaluation of GRASP against other interventions.		