

Outcomes of psychological therapies in prisons: a systematic review and meta-analysis

Isabel A. Yoon

Department of Psychiatry & Green Templeton College

University of Oxford

Submitted in fulfilment of the requirements for the degree:

Master of Science by Research in Psychiatry

Trinity Term, 2015

Word count: 10,811 words

General Abstract

Title: Outcomes of psychological therapies in prisons: a systematic review and meta-analysis

Name: Isabel Angela Yoon

Affiliations: Department of Psychiatry & Green Templeton College, University of Oxford

Degree: Master of Science by Research in Psychiatry

Term: Trinity Term, 2015

Abstract: ***Background:*** The imprisoned population, currently more than 10 million worldwide, is reported to have higher rates of suicide and prevalence of severe mental illness than the general population. With these identified needs, the effectiveness of psychological therapies on mental health outcomes in prisons warrants a systematic evaluation. The purpose of this thesis is to systematically review and analyse randomised controlled trials of psychological therapies in prisons.

Methods: A systematic search was conducted to identify studies that report mental health outcomes of psychological therapies among prisoners. Effect sizes were calculated from extracted data and presented using forest plots. Eligible studies were assessed for quality. Meta-regression analysis was performed to examine sources of heterogeneity between studies.

Results: Thirty-seven randomised controlled trial studies from 31 individual papers were eligible for data synthesis. In total the included studies showed a medium pooled effect size (0.50, 95% CI: 0.34-0.66) with a high level of heterogeneity. Subset of studies that adopted no treatment or waitlist as the control group had significantly larger pooled effect size (0.74, 95% CI: 0.51-0.97) than those that had treatment-as-usual (TAU) or another type of psychological therapy as controls (0.21, 95% CI: 0.01-0.41). Few studies that measured follow-up outcomes did not produce a significant pooled effect size. Effect sizes did not significantly differ between group therapy and individual therapy. Differences among varied treatment types were also not significant. Meta-regression of multiple study characteristics showed that presence of an adequate fidelity measure in a study was associated with lower effect size in treatment outcome. In the subgroup analysis with studies with no treatment/waitlist controls, lower participation rate and lower sample size were associated with higher effect size. Qualitative synthesis reported the difficulties in conducting RCTs of psychological therapy in prisons.

Conclusions: While psychological therapies in prisons tend to show medium treatment effect compared to no treatment or waitlist controls, the effect is reduced when the comparator is treatment-as-usual or other forms of psychological therapy. Studies should adopt an adequate fidelity measure, and caution is needed in interpreting studies with small sample sizes ($\leq$ 100) and high attrition rates ($\geq$ 20%).

Acknowledgements

I would like to thank everyone who contributed to the work reported in this thesis. I especially thank my supervisor Professor Seena Fazel for his guidance and constructive feedback throughout my research. I also thank Dr. Ram Benning for being the second data extractor and Dr. Pim Cuijpers for his advice on data analysis. I would like to thank Dr. Elizabeth Naomi Smith and Ms. Zuzanna Fimińska for their advice on the preparation of this thesis, and I am thankful to Ms. Saemi Kim for proofreading the text. I am grateful to Professor Willem Kuyken and Professor Klaus Ebmeier for their helpful feedback during my transfer of status examination. I am grateful to the Daehang Foundation of Hanmaum Seon Centre for providing the scholarship that made this research opportunity possible. I thank Dr. Miguel Farias, Dr. Xi Jing Chen, and Dr. David Gussak for providing the additional data of their studies for this review. Lastly, I would like to thank my family for their support as always.

Table of Contents

General Abstract ii
Acknowledgements iii
Table of Contents iv
Abbreviations and symbols used in figures and tables vi
Index of Figures vii
Index of Tables viii
Chapter 1: Introduction 1
1.1 Background 1
1.2 Aims 3
Chapter 2: Methods 4
2.1 Search strategy 4
2.2 Study eligibility 5
2.3 Data extraction and quality assessment 7
2.4 Statistical analysis 8
2.4.1 Effect size calculation 8
2.4.2 Meta-analysis 9
2.4.3 Meta-regression & publication bias 11
Chapter 3: Results 14
3.1 Main results 14
3.1.1 Study characteristics 14
3.1.2 Types of outcome measures 15
3.1.3 Treatment length & quality rating 16
3.1.4 Effect sizes for total studies 17
3.1.5 Specific types of outcomes 18
3.1.6 Follow-up outcomes 19
3.2 Secondary & subgroup analyses 19
3.3 Meta-regression results 22
3.3.1 Univariate meta-regression analysis 22
3.3.2 Multivariate meta-regression analysis 23
3.3.3 Exploratory subgroup analysis 24
3.4 Difficulties of conducting RCTs in prisons (qualitative analysis) 24
3.5 Publication bias 27

Chapter 4: Discussion 28
4.1 Overview 28
4.2 Principal findings 28
4.3 Meta-regression analysis results 32
4.4 Treatment length 33
4.5 Difficulties of conducting RCTs in prisons 34
4.6 Publication bias 34
4.7 Strengths & limitations 35
4.7.1 Overview of strengths 35
4.7.2 Study quality & characteristics 35
4.7.3 Methodology of research – search terms 38
4.7.4 Effect size calculation 39
4.7.5 Variation in outcome types 40
4.7.6 Heterogeneity in studies 41
4.8 Clinical significance of findings 41
4.9 Conclusion 42
Appendix A - PRISMA Checklist 47
Appendix B – RCT Quality checklist used for quality appraisal 50
Appendix C – Study characteristics of included studies 52
Appendix D – Quality rating of included studies (using quality checklist from Appendix B) 59
Appendix E – Effect sizes of studies with depression outcomes (by comparator type). 63
Appendix F – Effect sizes of studies with anxiety outcomes (by comparator type). 64
Appendix G – Effect sizes of studies with overall psychopathology outcomes (by comparator type) 65
Appendix H – Effect sizes of studies with trauma outcomes (by comparator type). 66
Appendix I – Contoured funnel plot of total studies. 67
Appendix J – Contoured funnel plot of studies with no treatment/waitlist controls 68

Abbreviations and symbols used in figures and tables

- MDD = Major Depressive Disorder
- SUD = Substance Use Disorder
- PTSD = Post Traumatic Stress Disorder
- CI = confidence interval
- TR = treatment
- TAU = treatment-as-usual
- ES = effect size (standardised mean difference)
- BDI = Beck Depression Inventory
- STAI = State-Trait Anxiety Inventory
- HADS = Hospital Anxiety and Depression Scale
- SCL-90 = Symptom Checklist-90
- SCL-90-R = Symptom Checklist-90-Revised
- β = coefficient
- s.e. = standard error
- p = p-value of t-test

Index of Figures

Figure 1 – PRISMA flow diagram 14
Figure 2 – Effect sizes of included studies stratified by comparator type 18
Figure 3 – Effect sizes of total studies stratified by treatment type. 22

Index of Tables

Table 1 − Inclusion and exclusion criteria for data synthesis 7
Table 2 – Effect sizes of included studies stratified by participation type in therapy..... 20
Table 3 – Effect sizes of included studies stratified by treatment type..... 21
Table 4 – Univariate analysis results of total studies 23

Chapter 1: Introduction

1.1 Background

There are currently more than 10 million people in prisons worldwide.[1] Mental health problems are common in prisoners: the 6-month prevalence of major depression was 11% and that of psychosis was 4% among prisoners.[2] In the general population the 12-month prevalence of major depression was 6%[3] while that of psychosis was 1%.[4] Reviews have also estimated that prevalence of anxiety disorders, post-traumatic stress disorder, and personality disorder are also higher than that of the general population.[5-8]. Psychological and psychiatric problems are thought to be major risk factors for a range of adverse outcomes in prison and also on release. These include suicide, estimated to be five-fold higher in men inside prison and around 20-fold higher in women in England and Wales.[9,10] It is thought that one of the main reasons for increased suicide rates in prison may be insufficient psychological treatment, based on the fact that about half of the prisoners that die from suicide have a treatable mental illness.[11] Higher rate among teenagers can be due to their higher risk of having an undetected mental illness and untreated substance abuse problems.[11]

Moreover, self-harm is reported to be 20–24% in female prisoners and 5–6% in male prisoners in any given year.[12] Risk factors for such suicide and self-harm include mental health problems.[13,14] In addition, it is thought that reoffending risk is related to mental health conditions, particularly if they are untreated.[15,16]

Treatments for mental health conditions in prisons are inconsistent with some prisons offering psychological therapies and some that do not, and many that focus solely on pharmacological treatments for substance abuse and severe mental illness. However, little

is known about what treatments have the best evidence base in prisons, and in particular, with an increasing number of psychological therapies being offered, it is important to clarify the current status of treatment in prisons. Given the fact that prisoners are restricted from using the treatment resources outside prisons, it is important to examine the effectiveness of psychological therapies that are being provided inside the correctional facilities.

Previously, there have been a number of systematic reviews of interventions for mental health in prisoners.[17-25] Of these, one systematic review looked at interventions for anxiety, depression, and post-traumatic stress disorder (PTSD) in adult offenders.[23] However, this review included pharmacological treatments and 11 psychological trials with only adult offenders. Another review focused on non-pharmacological interventions with women offenders.[17] However, this review was limited by the fact that it included non-randomised controlled trials (non-RCTs) and meta-analyzed only subgroups with 5 or less studies. Other reviews have focused on one specific disorder (PTSD),[20] specific outcome (violent behaviours),[25,26] and narrative reviews of one specific therapy[21] or intervention type.[27] One review examined group therapy with incarcerated offenders and included non-RCTs.[24] Other reviews have been broader literature reviews that explored many different types of papers on mental health services for prisoners (including theoretical papers, audits, needs assessments, and screenings for mental health problems)[18] or a review of interventions for adult offenders with serious mental illness that included interventions outside prison and pharmacological treatments as well.[19] Another recent review covered RCTs of interventions to improve health during imprisonment and a year after release, but this review covered a wide range of mostly physical health and drug abuse interventions,[22] and it did not include a meta-analysis.

Thus, many of the previous reviews are based on few studies and focused on specific disorders or intervention types instead of including all psychological therapies in prisons. Moreover, none of the previous reviews have examined only RCTs of psychological therapies excluding pharmacological interventions to my knowledge.

This thesis attempts to overcome such existing gaps in research, by conducting a systematic review and meta-analysis on the RCTs of psychological therapies in prisons using unselected samples for all mental health conditions across gender and age groups.

The method of systematic review was chosen as it allows for an impartial assessment of evidence through a systematic approach, which minimizes biases and systematic errors.[28] A meta-analysis is a statistical method that derives an estimated mean effect by combining estimates of the effect sizes from individual studies.[29]

Only randomised controlled trials (RCTs) were included in this review since meta-analyses can produce the most unbiased effect estimate when conducted exclusively on reports of RCTs.[30] Because of randomization, RCTs are free from the effect of various unmeasured confounders, unlike observational studies that are at high risk of bias and confounding.[31]

1.2 Aims

The aim of this research is to systematically review outcomes of psychological therapies in prisons. Specifically, this review aimed to compare effect sizes across different types of psychological therapies and examine sources of heterogeneity. It also aimed to qualitatively analyse the included studies by examining the difficulties in implementing RCTs of psychological therapies in prisons.

Chapter 2: Methods

2.1 Search strategy

The following databases were searched from their start dates until 30th May, 2015: PsycINFO, MEDLINE, Global Health, PubMed, CINAHL, National Criminal Justice Reference Service, Scopus, EMBASE, and Cochrane Library. National Criminal Justice Reference Service and CINAHL had to be searched on a separate portal from rest of the databases in order to use the search terms with wildcard options. PubMed, Scopus, EMBASE, and Cochrane Library were also searched separately from the portal that combined PsychINFO, MEDLINE, and Global Health. Additional targeted searches were conducted by hand-searching citations and reference lists of other systematic reviews and articles. Targeted searches on specific authors (identified from previous papers), studies on mindfulness-based therapies, and the treatment of psychopathy in prison were also conducted separately. Authors were contacted when more data were needed.[32-36]

The following keywords were used to identify psychological disorders: mental*, psych*, disorder*, depress*, schizo*. These terms were combined with search terms relating to the prison setting (prison*, inmate*, jail*, penal, correctional, incarcerated) and intervention (cognitive behavioral therapy, cognitive behavioural therapy, CBT, group therap*, intervention*, treatment*, therapeutic communit*) and study design (RCT*, controlled trial*, randomi#ed controlled trial*, controlled clinical trial*, randomi#ed trial*, randomi#ed clinical trial*, trial*). Additional targeted search for psychological therapies that may have been missed by other search terms were performed with the keywords: therap* and psychotherapy.

The guidelines in *The Cochrane Handbook* were followed for specific methods for the search strategy as well as the analysis and summary of data.[28] Additionally, 'Preferred Reporting Items for Systematic Reviews and Meta-analyses' (PRISMA) guidelines were followed (Appendix A).[37] PRSIMA guidelines are an update on the guidance outlined in the QUOROM Statement (Quality of Reporting Meta-analyses), which was developed in 1996.[37] The PRISMA Statement consists of a 27-item checklist and a four-phase flow diagram.

2.2 Study eligibility

This review investigated studies that: (a) are randomised controlled trials of psychological therapies in prisons; (b) compare psychological therapy with control groups, which are treatment-as-usual (TAU), waitlist, or another type of psychological therapy; (c) report a standardised outcome measure of mental health.

The reason for excluding other offenders was that this review aims to examine the efficacy of psychological therapies in prisons and not in other specialised settings (e.g. forensic psychiatric hospitals outside prisons) or in the community (e.g. post-release treatments).[38,39] In order to minimize the bias of non-random group assignment, studies that were not randomised (including pre-test/post-test comparisons with no control group) were excluded.[40-49] For example, one study implemented random assignment of participants to each group but not a random selection of the total sample (treatment was mandated for the participants who were prisoners with substance abuse issues), and thus was excluded.[35] Studies that only reported outcomes on substance abuse or recidivism were also excluded.[50,51] Studies on the treatment of psychopathy or sociopathy in prisons were not included because none of the identified studies had standardised psychological outcome measures required for this review.[52-56] Other studies were also excluded if they did not include standardised psychological measures.[57]

One study reported on the outcomes of a prison therapeutic community at 6-months and 12-months since prison release.[36] This study was excluded because it compared the treatment outcome with control group participants that received intensive outpatient treatment; therefore, this was not comparable to other included studies. One study was excluded because it only recorded psychological measures (Beck Depression Inventory) of the treatment group and not the control group, even though it had substance use outcomes for the control group as well.[58] When pre-test and post-test scores were compared to calculate an effect size, it resulted a significantly inflated effect size (near 4.0) compared to those calculated from control vs. treatment groups of other studies. Details of the inclusion and exclusion criteria are summarised in Table 1.

Inclusion Criteria	**Exclusion Criteria**
Design: Randomised controlled trials including pilot studies and cluster-randomised trials	Non-randomised trials (including pre-test/post-test comparisons), narrative reviews, and case studies
Population: Prisoners (including youth of ages 12–25)[1]	Populations not currently in prison/correctional settings (e.g. post-prison release treatments, offenders on parole, forensic psychiatric hospitals and therapeutic communities outside prisons)
Interventions: Cognitive Behavioural Therapy, Dialectical Behaviour Therapy, Mindfulness-based Therapy, other group therapies such as Music Therapy and Art Therapy (including self-help treatments)	Interventions: Medication only
Outcomes: Psychological improvement	Outcome: Recidivism, substance use

[1] Definition of youth varied across institutions: Four UK Secure Children Homes (SCH) and a Youth Offender Institution (YOI) housed ages 13-18, a Scotland YOI housed ages 16-21, one USA Youth Correctional Facility (YCF) housed ages 15- 18, and USA Oak Creek YCF housed ages 12-25.

measured by standardised measurements (e.g. Beck Depression Inventory) Post-treatment measures, some studies including follow-up measures (3-month and 6-month)	
Studies in any language and published at any date prior to the last date of the search (30th May 2015). Unpublished studies, such as doctoral theses, were also considered.	Studies that did not provide available data to calculate effect sizes, even after contact with the authors, were excluded.

Table 1 - Inclusion and exclusion criteria for data synthesis

2.3 Data extraction and quality assessment

In addition to effect sizes, 95% confidence intervals, variance of the primary outcome and secondary outcomes, and the following background characteristics were recorded in the data extraction to be explored as sources of heterogeneity: year of publication, location, sample size, percentage of male, exclusion criteria, diagnosis (if reported), type of treatment, and quality rating. Primary outcome was selected on the basis of most commonly used psychological assessment in the included studies so that studies could be directly compared. All reported difficulties in implementing RCTs in prison were recorded. A second extractor (a specialist psychiatrist, Dr Ram Benning, who has worked in prison health) crosschecked extracted data by working independently, and any discrepancies were resolved by discussion between the extractors.

Eligible studies were assessed for quality using the quality checklist for an RCT from a National Clinical Practice Guideline commissioned by the National Institute of Health and Care Excellence (NICE) (Appendix B).[59] The checklist included items to assess internal validity such as the use of adequate concealment method for participant allocation, blinding of subjects and investigators, and the use of intent-to-treat analysis that analyses all subjects randomly allocated to each group, regardless of whether they have completed

the treatment. Overall rating was either − (few or no criteria fulfilled), + (some fulfilled), or ++ (all or most fulfilled).

This review was registered on the international database of prospectively registered systematic reviews in health and social care (PROSPERO) in January, 2015.[60]

2.4 Statistical analysis

2.4.1 Effect size calculation

Standardised Mean Difference (d), 95% confidence intervals, and variance were calculated for each study with the mean and standard deviations of the psychological scores of the experiment and control groups. Calculations were done through the "Practical Meta-analysis Effect Size Calculator" available on the Campbell Collaboration website.[61] For studies that reported standard errors instead of standard deviations, a different function of the effect size calculator was used. Likewise was done for studies that reported p-values of t-tests instead of standard deviations.[62] For effect size and confidence intervals, negative signs were reversed in order to follow the convention of having positive effect sizes for positive treatment effect. In the raw scores of psychological outcome, lower scores indicated psychological improvement since scores were higher with greater severity or number of symptoms. For a study that reported only the change scores[63] or scores that were higher with better health outcome (including secondary outcomes),[62,64-66] the signs were not converted. For studies that had more than one type of control group, the more treated control group was chosen over waitlist control for the effect size calculation to have a more conservative estimate of the treatment effects.[65,67] For a paper that compared two different treatment groups (three groups in one study), each treatment group was independently compared with the control group.[63,68] However, for one study with two treatment groups (Male-Trauma Recovery Empowerment Model: M-TREM and Seeking Safety) 3-month and 6-month follow-up data were only available for the treatment groups

and not the control group.[64] Thus, to fully utilize the available data, effect sizes for follow-up time points were calculated from the comparison of the two treatment groups to each other. The immediate post-treatment effect size was calculated by comparing each treatment group with the control, as done with other cases.

2.4.2 Meta-analysis

The random-effects model was used for the meta-analyses of effect sizes; this model assumes that the included studies are estimating different, but related, intervention effects.[69] Compared to the fixed-effect model that assumes that the true effect size estimated by each study is exactly the same, a random-effects model assumes that the true effect size is different in each study and that the pooled studies represent a random sample of all observations.[28,69,70] Specifically, the estimation of the intervention effects is expected to follow a distribution across the studies, and study weights are adjusted according to the extent of the between-study variation (heterogeneity).[28] Therefore, it produces wider confidence intervals when there is heterogeneity.[69] Moreover, the random-effects model assigns similar weights to all studies included in the meta-analysis regardless of sample size, so the overall estimate is less likely to be biased by a few large studies.[69] Given the clinical differences between the studies such as different types of psychological therapies, countries, and age group of the sample, the random-effects model was adopted.

Levels of heterogeneity was assessed using I^2, which represents the percentage of the observed variation in effect size across studies due to true heterogeneity rather than chance.[71] This measure is preferred over the Q statistic (a measure of weighted sum of squared standard deviations of each study's effect estimate from the overall effect estimate) and tau-squared (estimate of between-studies variance in true effect size) because it is an intuitive measure that is independent of the number of studies included and the

metric of the effect size.[69,71,72] Suggested benchmarks for I^2 are that values of 25%, 50%, and 75% could indicate low, moderate, and high levels of heterogeneity, respectively.[73]

Effect sizes were grouped into domains and combined on forest plots. First, the studies were grouped by comparator type: first group of studies included no treatment, no-contact group, and waitlist as controls. Second group included treatment-as-usual or another type of more intensive control (condition that asked a group to write about "trivial topics" in the "writing about trauma" therapy[67]). Third group included studies with controls as another form of psychological therapy such as individual supportive therapy, standard prison-based therapeutic community, supportive group therapy (SGT) or attention-matched manualised psychoeducation (level of attention required matched to treatment group).[66,74-77]

Second, the studies were stratified by treatment type. Treatment types were categorised into CBT-based therapies, mindfulness-based therapies, trauma-based therapies, and others. These categories were designed to include at least 5 studies (the minimum pre-specified number of studies for analysis in this review) in each category when examining all of the included studies together. In the CBT-based category, therapies that used traditional CBT techniques and new types of therapies using the principles of CBT (e.g. Seeking Safety[64,78,79] and Cognitive Processing Therapy[80]) were included. For the mindfulness-based category, treatments that used meditation were included (e.g. Transcendental Meditation[81]). In the trauma-based category, treatments that focused on treating trauma symptoms were included. One example in this category is Trauma Incident Reduction Therapy, which focuses on the memory instead of the current symptoms because "current symptoms are believed to be related to a past event."[82]

In the "other" category various therapies such as Interpersonal Psychotherapy (IPT),[75] mother-infant attachment based therapy,[83] Gender Responsive Treatment (GRT),[77] as well

as art therapy[34,84] and music therapy[32,85] were included. Two therapies that combined principles of both CBT and mindfulness (Acceptance & Commitment Therapy (ACT)[68] and Gender Responsive Treatment (GRT)[77]) were placed into the other category instead of either category of CBT or mindfulness. Test of between-group heterogeneity of these subgroups by treatment type were conducted using the mixed-effect analysis, which adopts random effects model within subgroups and fixed effect model across subgroups.[69]

In further analysis, the studies were grouped by type of participation in therapy: group, individual, or combination that included both group and individual settings in the therapy. Separate analyses were done with results that measured one specific type of psychological outcome (e.g. depression, anxiety, trauma symptoms, and overall psychopathology). There were some other outcome measures that were of possible interest, including measures on coping skills[64-66,86] and hostility/anger.[66,81,87] However, there were less than five studies meeting these criteria, so these were not examined in the meta-analysis. Other outcome measures that seemed to be at least partially overlapping with the primary outcome measures (e.g. hopelessness scale with depression scale) were also not analysed separately. Even though the main analysis was done with psychological outcomes, a self-reported measure of physical health (including sleeping difficulties[62]) was analysed in an exploratory analysis, as it was available in five studies.

2.4.3 Meta-regression & publication bias

Meta-regression analysis of various study characteristics was performed to examine sources of heterogeneity. A meta-regression tests for the presence of a linear relationship between the intervention effect and one or more explanatory variables, and the coefficient value of the analysis quantifies the amount of change in the intervention effect with a unit increase in the explanatory variable (the potential effect modifier).[69] Studies of higher weight have more influence on the relationship than those of lower weight.

Pre-specified factors examined include gender, mean age (continuous variable), age group (adult v. adolescent), year of publication, country (USA v. rest of the world; 26 studies out of 37 studies were from the U.S.), attrition rate (both continuous and < 20% v. ≥ 20%), sample size (both continuous and ≤ 100 v. > 100), diagnosis (required v. not required), and study quality (high v. medium). The variables treatment length (in unit of weeks), treatment fidelity measure (recorded as 1 = adequate measure present, 0.5 = some measure present yet not fully satisfactory, and 0 = absence), and type of control group (no treatment/waitlist v. TAU/other therapy) were added for analyses based on the feedback from examiners of an earlier presentation of the findings. For the dichotomous version of the gender variable, more than 90% of male was classified as male even when total sample included some females. In addition, studies that did not explicitly state drop-out rates were excluded in the analysis with attrition rates.[62,80,88]

Each variable was first examined in a univariate analysis individually and later explored together in a multivariate analysis. Multivariate analysis was conducted with all of the variables together. This analysis was done with each combination of variables to include only one of the two versions of the same characteristics (i.e. in the case of gender, age, sample size and participation rate) in order to avoid the problem of collinearity between variables, which can complicate the results of variables that are highly dependent to one another.[89] If there were less than 10 studies that reported the variable, analysis was not performed as statistical power was limited.[90] Therefore, in the subgroup analyses multivariate analyses were conducted with just the variables that had a p-value of less than 0.1 from univariate analysis.

To test for publication bias, funnel plot analysis and Egger's test were performed. A funnel plot is a scatter plot of effect estimates from individual studies against each study's measure of precision, which in this review was the standard error.[91-93] Egger's test can

reveal asymmetry in the funnel plot by using a regression method.[93] Other tests for publication bias such as Begg's test and Harbord's test were not used, as they are only applicable for dichotomous or ordinal outcomes.[28,92] As an exploratory analysis, the trim and fill analysis (with random-effects model) was also conducted with the total sample and subset of samples (studies with no treatment/waitlist controls) to identify and correct for funnel plot asymmetry due to publication bias.[28] This method eliminates smaller studies that cause the funnel plot asymmetry, estimate the true centre of the funnel with the trimmed funnel plot, and substitutes the removed studies with their missing estimated 'counterparts' around the centre.[28,94] However, findings from this analysis must be interpreted with caution, especially if there is substantial between-study heterogeneity.[94] This method assumes that there should be a symmetric funnel plot in absence of publication bias, but asymmetry in the funnel plot can arise from other sources of heterogeneity in the studies.[94]

All analyses were performed in Stata statistical software package (STATA-IC) version 14 (Stata Corp, 2015) using the following commands: metan (for random-effects meta-analysis), metareg (for meta-regression), metabias (for publication bias analysis), metatrim (for trim-and-fill analysis), and heterogi (for calculation of confidence intervals for heterogeneity level).

Chapter 3: Results

3.1 Main results

3.1.1 Study characteristics

From a combination of 1,430 records initially found on separate database search portals and additional targeted searches, 688 records were screened by title or abstract (Figure 1). Full texts of 71 articles were screened for eligibility, and 37 studies from 31 individual papers were included in the quantitative and qualitative analysis.[32-34,62-68,74-88,95-100]

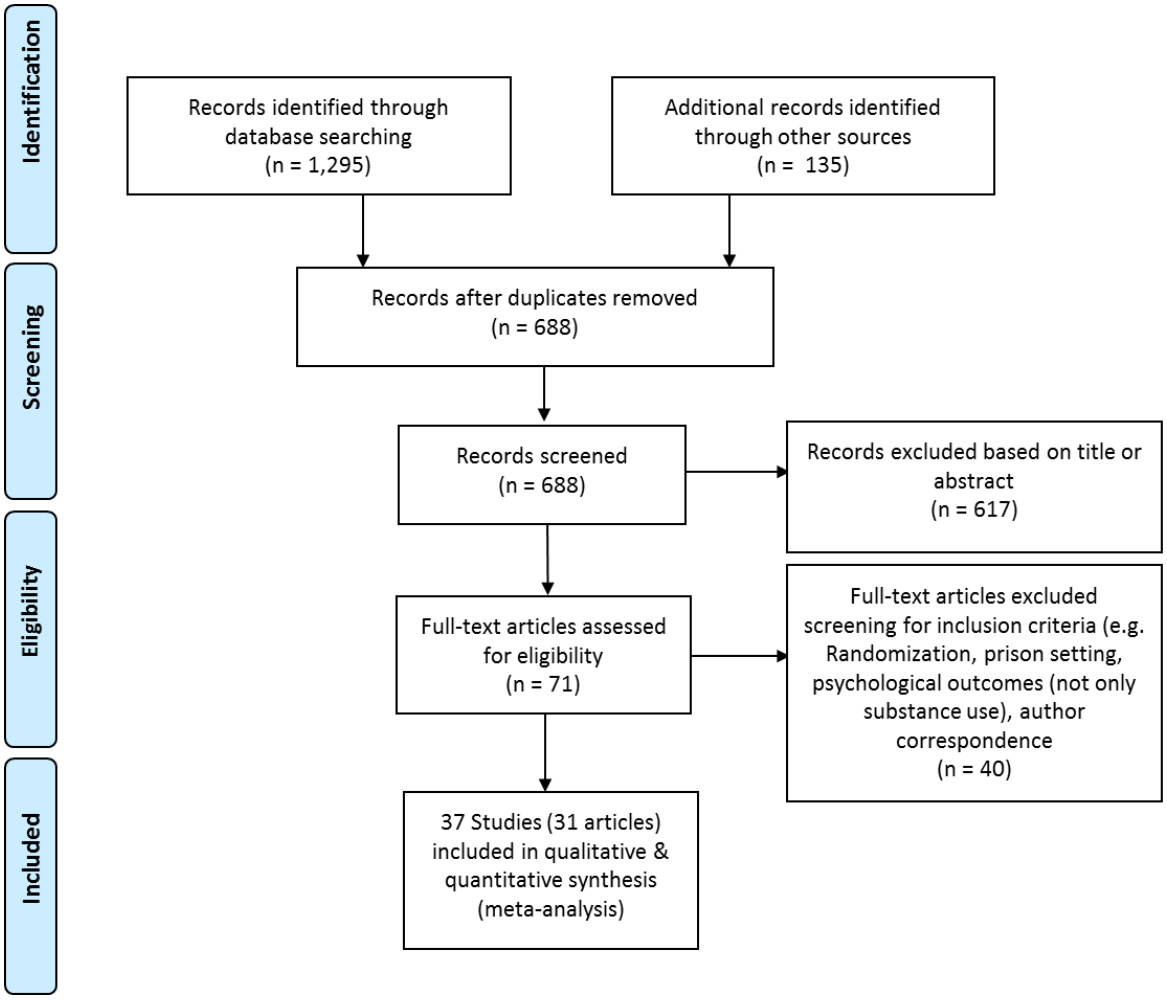


Figure 1 – PRISMA flow diagram

The included studies were published between 1979 and 2015 and were from 7 different countries including the UK, the United States, China, Iran, Norway, India, and Spain. The total sample had 2,761 prisoners, 59% of which were male (except for 2 studies, all had either exclusively male or female samples). The mean age of the total sample was 31.8 years (adult sample: 34.4 years, juvenile sample: 16.9 years).[65,80,86,95] All of the studies recruited voluntary participants.

All studies reported immediate post-treatment outcomes with 6 studies reporting 3-month post-treatment follow-up outcomes and 5 studies reporting 6-month follow-up outcomes. Most of the studies did not require a diagnosis for the participants to be in the study, but 16 of the studies had a specific diagnosis or symptoms in their participant inclusion criteria, namely depression, anxiety, PTSD, or substance use disorder. A complete list of study characteristics of the included studies is found in Appendix C.

3.1.2 Types of outcome measures

The primary outcome measures reported in the study include: Beck Depression Inventory (BDI),[101] BDI-II,[102] Hamilton Rating Scale for Depression (HRSD),[103] global severity index of the Brief Symptom Inventory (BSI),[104] Clinician Administered PTSD Scale (CAPS),[105] State-Trait Anxiety Inventory (STAI),[106] Youth Self Report (YSR),[107] global severity index of the Symptom Checklist Revised (SCL-90-R),[108] Cognitive Somatic Anxiety Questionnaire (CSAQ),[109] psychological composite score of the Addiction Severity Index (ASI),[110] Hospital Anxiety and Depression Scale (HADS),[111] Center for Epidemiologic Studies Depression Scale (CES-D),[112] ASI (Anxiety Sensitivity Index) total score,[113] and sleeping difficulties (a scale from 0 to 4 on the degree to which sleeping difficulties bothered participants).[62] Except for the sleeping difficulties measurement used in the meditation study by Sumter et al. (2009),[62] all of the primary outcome measures

used in the main analysis were standardised and validated outcome measurements with calculated reliability scores.

Other outcome measures used for the analysis of specific types of psychological outcomes include the following: for depression, depression subscale of Symptom Checklist (SCL-90-R),[108] Hospital Anxiety and Depression Scale (HADS),[111] State-Trait Anxiety Inventory (STAI),[106] and a scale for emotions of hopelessness[62] were used. For anxiety, Hamilton Anxiety Rating Scale (HARS),[114] Clinical Anxiety Scale (CAS),[115] tension reduction behaviour subscale of Trauma Symptom Inventory (TSI),[116] anxious arousal subscale of TSI, State Anxiety scale of STAI, anxiety subscale of SCL-90-R, anxiety subscale of HADS, Perceived Stress Scale (PSS),[117] and anxiety subscale of BSI were used. For overall psychopathology, Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM),[118] positive symptom score of BSI, and internalizing score of YSR were used. For measuring PTSD outcomes, PTSD Symptom Scale-Self Report (PSS-SR),[119] total PTSD symptoms score of CAPS, and intrusive experiences subscale of TSI were used. For an exploratory analysis in physical health measures, physical health composite score of ASI-6 (Addiction Severity Index-6),[120] Coping Resources Inventory for Stress (CRIS)-physical health subscore, and the number of post-test infirmary visits[67] were assessed.

3.1.3 Treatment length & quality rating

Treatment length ranged from a brief one-time treatment (3-4 hour)[82] to longer treatments up to 18 weeks[87] (except one program for a prison therapeutic community that was minimum 6 months).[77] Average length of treatment for all studies was 10 weeks. Quality ratings of all included studies were either ++ (all or most criteria fulfilled) or + (some of the criteria fulfilled) (Appendix D). Because there were no studies rated as − (none or few

criteria fulfilled), none of the studies were excluded based on quality rating. Only 7 studies out of the 37 studies included met the highest quality rating.[75-77,82,83,86,100]

In addition, there were 12 studies with a satisfactory fidelity measure of treatment, [64,76,78,85-87,95-97,100] 5 studies with some measure of fidelity measure (not fully satisfactory),[65,68,75,98] and 9 studies without any measure.[63,66,74,77,80,82,83,100] Remaining 11 studies did not mention the presence or absence of a fidelity measure.[32-34,62,67,81,84,88,99] An example of a satisfactory measure include fidelity assessment in each session of treatment by the co-therapist according to detailed manuals and fidelity checklists.[76] Suboptimal fidelity included weekly team meetings of the treatment providers to discuss the procedure and progress of the treatments.[68]

3.1.4 Effect sizes for total studies

In total, the included studies had a medium pooled effect size according to Cohen's standards[121] (ES = 0.50, 95% CI: 0.34–0.66) with moderate to high levels of heterogeneity (I^2 = 73%, 95% CI: 62%–80%).[71] When studies were divided into three different subgroups by comparator type, the effect size was larger in the subgroup with no treatment/waitlist controls (ES = 0.74, 95% CI: 0.51–0.97) than in the subgroup with TAU controls (ES = 0.19, 95% CI: -0.07–0.45) (Figure 2). Studies with another psychological therapy as controls also had a smaller effect size than those with no treatment/waitlist controls, although there was slight overlap in the confidence intervals (ES = 0.25, 95% CI: -0.09–0.59). Heterogeneity was moderate to high in all conditions (I^2 = 74%, 95% CI: 60%–83%; I^2 = 68%, 95% CI: 40%–83%; and I^2 = 52%, 95% CI: 0%–83%, respectively). Even when studies with TAU or other psychological therapy controls were combined, they had a significantly lower effect size than the no treatment/waitlist control studies (ES = 0.21, 95% CI: 0.01–0.42; I^2 = 63%, 95% CI: 36%–78%).

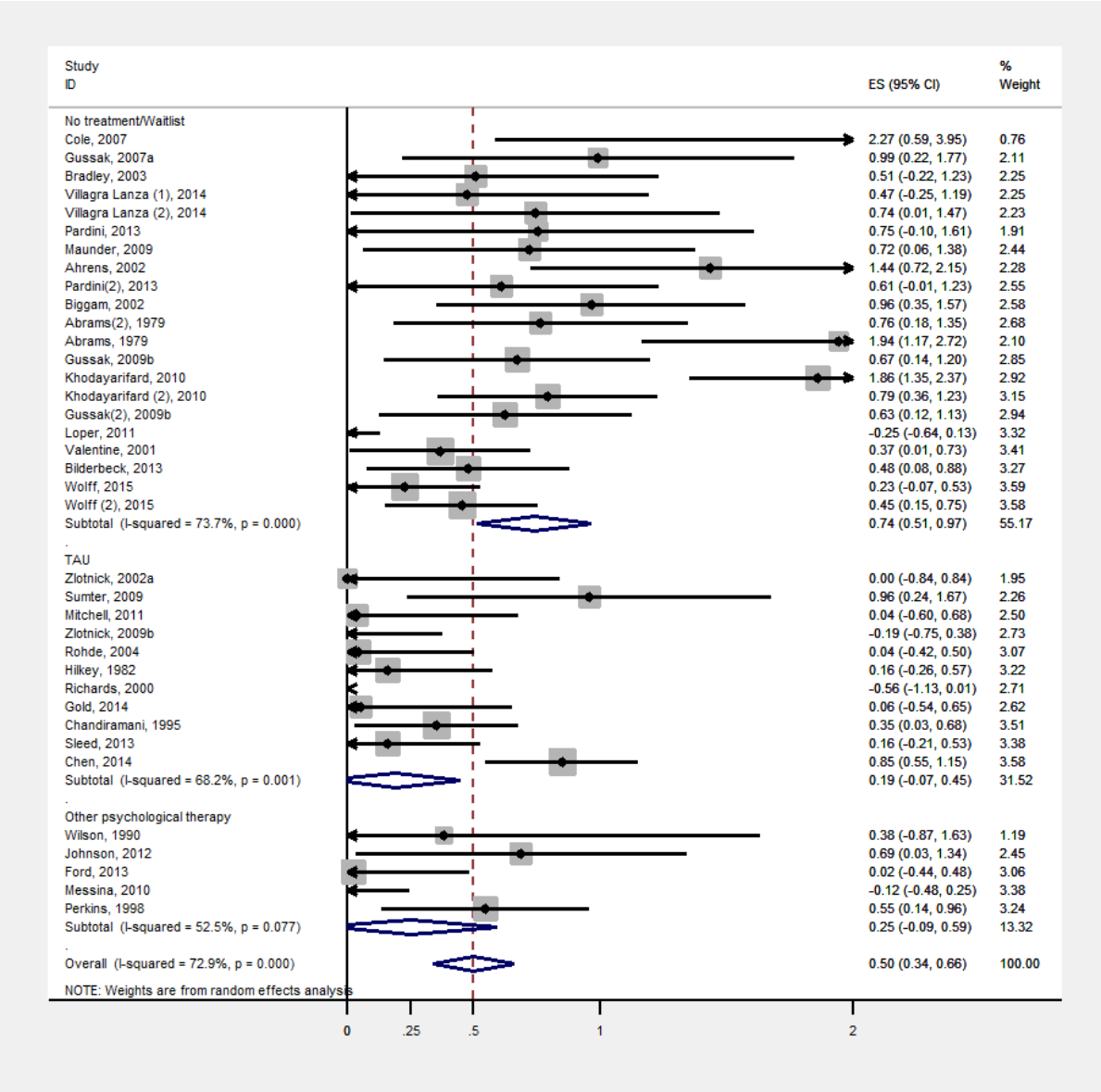


Figure 2 – Effect sizes of included studies stratified by comparator type

3.1.5 Specific types of outcomes

In the analyses of specific types of outcomes, 20 studies that measured depression outcomes produced a pooled effect size of 0.60 (95% CI: 0.38–0.83) with moderate to high heterogeneity (I^2 = 71%, 95% CI: 54%–81%) (Appendix E). Twenty studies with anxiety outcomes produced a pooled effect size of 0.57 (95% CI: 0.31–0.84) with high heterogeneity (I^2 = 83%, 95% CI: 75%–89%) (Appendix F). Seventeen studies with outcome measures of overall psychopathology produced a pooled effect size of 0.32 (95% CI: 0.05–0.59) with high heterogeneity (I^2 = 81%, 95% CI: 70%–88%) (Appendix G). Ten

studies with trauma outcomes had a pooled effect size of 0.35 (95% CI 0.14–0.56; I^2 = 44%, 95% CI: 0%–73%) (Appendix H). Lastly, 6 studies with physical health outcomes had non-significant effect size of 0.18 (95% CI: -0.32–0.68; I^2 = 72%, 95% CI: 35%–88%).

3.1.6 Follow-up outcomes

Six studies that measured follow-up outcomes at 3 months after the end of treatment had a non-significant pooled effect size of 0.29 (95% CI: -0.05–0.64; I^2 = 62%, 95% CI: 8%–84%).[64,75,78,79,82,95] One study with the largest sample that compared two treatment types to each other instead of a control was taken out in a sensitivity analysis: the result was still not significant (ES = 0.35, 95% CI: -0.09–0.79).[64] Similarly, 5 studies (including two studies from one paper) that measured follow-up outcome at 6 months after treatment had a non-significant pooled effect size of 0.06 (95% CI: -0.15–0.26; I^2 = 0%, 95% CI: 0%–79%).[64,68,77,79]

There were studies with outcomes at different time points such as 6-week post-treatment,[67] 1-year post treatment,[77] and mid-test,[32,74] but these data were not analysed as there were less than five studies providing measures at these time points.

Among specific types of outcome measures, only the measure of overall psychopathology had at least five studies that recorded follow-up.[64,68,77,79] At 6 months after the end of treatment, the effect size was not significant (0.05, 95% CI: -0.18–0.28; I^2 =15%, 95% CI: 0%–82%).

3.2 Secondary & subgroup analyses

When 21 studies with no treatment/waitlist controls were analysed separately, group therapy and individual therapy showed almost no difference (ES = 0.59, 95% CI: 0.37–0.81; ES = 0.59, 95% CI: 0.36–0.81, respectively). Combination therapy had a larger pooled effect size, but the difference was not significant from the other two types (ES =

0.94, 95% CI: 0.29–1.58). Compared to individual therapies (I^2 = 0%, 95% CI: 0%–79%) and group therapies (I^2 = 35%, 95% CI: 0%–70%), combination therapies had a higher level of heterogeneity (I^2 = 90%, 95% CI: 82%–94%). The findings were similar across all of the studies and in the other subgroup with TAU/other therapy controls (Table 2). However, the second subgroup had less than 5 studies for each of the individual and combination types, so analyses for these subsets were limited.

Therapy type	Control group	Number of studies	ES	95% CI	I^2
Group	No TR/waitlist	9	0.59	[0.37, 0.81]	35%
	TAU/other	11	0.29	[0.06, 0.53]	63%
	Total	20	0.43	[0.26, 0.60]	55%
Individual	No TR/waitlist	5	0.59	[0.36, 0.81]	0%
	TAU/other	2	-0.28	[-0.86, 0.31]	47%
	Total	7	0.38	[0.02, 0.74]	65%
Combination	No TR/waitlist	7	0.94	[0.29, 1.58]	90%
	TAU/other	3	0.19	[-0.18, 0.56]	35%
	Total	10	0.72	[0.25, 1.19]	87%

Table 2 – Effect sizes of included studies stratified by participation type in therapy.

When the studies with no treatment/waitlist controls were stratified by treatment type, they did not significantly differ in effect size (Table 3). Rest of the studies with TAU/other therapy controls showed similar results. All of the 37 studies were again stratified by treatment type and findings were similar in that the difference between effect sizes of each treatment type was not statistically significant (Figure 3, Table 3).

Treatment type	Control group	Number of studies	ES	95% CI	I^2
CBT-based	No TR/waitlist	9	0.74	[0.29, 1.20]	85%

	TAU/other	5	-0.003	[-0.29, 0.28]	0%
	Total	14	0.50	[0.16, 0.85]	81%
Mindfulness-based	No TR/waitlist	2	1.32	[0.17, 2.48]	8%
	TAU/other	3	0.50	[0.23, 0.77]	16%
	Total	5	0.83	[0.38, 1.28]	73%
Trauma-based	No TR/waitlist	4	0.49	[0.17, 0.81]	36%
	TAU/other	2	-0.24	[-0.81, 0.32]	58%
	Total	6	0.27	[-0.12, 0.65]	70%
Others	No TR/waitlist	6	0.65	[0.43, 0.87]	0%
	TAU/other	6	0.30	[-0.05, 0.65]	76%
	Total	12	0.47	[0.24, 0.69]	61%

Table 3 – Effect sizes of included studies stratified by treatment type.

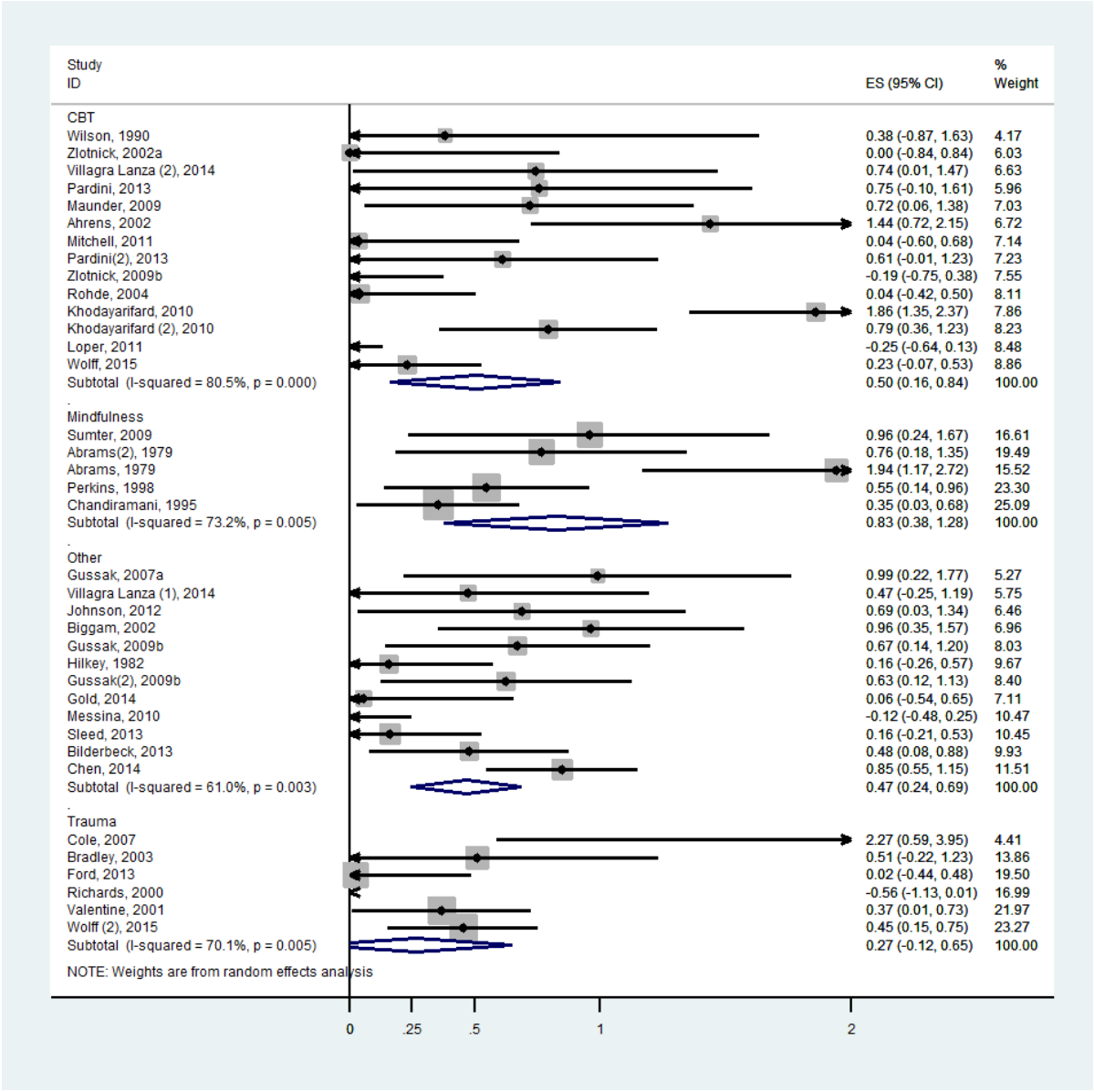


Figure 3 – Effect sizes of total studies stratified by treatment type.

3.3 Meta-regression results

3.3.1 Univariate meta-regression analysis

Variables for participation rate (both continuous and dichotomous) and control type (dichotomous) showed significant association ($p < 0.05$) with the effect size in the univariate analysis (Table 4). Coefficients of these variables suggested that greater attrition rates and no treatment/waitlist controls were associated with higher effect size.

Variable	β	s.e. (β)	p
Gender	0.2687	0.1839	0.153
Mean age (continuous)	0.0097	0.0127	0.450
Age group (Adult v. adolescent)	-0.0864	0.2970	0.773
Year of study (continuous)	-0.0085	0.0095	0.376
Country (USA v. rest of the world)	-0.2504	0.1905	0.197
Participation rate:			
Continuous (attrition rate)	1.5468	0.6212	**0.018***
Low ($\leq$ 80%) v. high (> 80%)	-0.4095	0.1751	**0.025***
Sample size			
Continuous	-0.0024	0.0017	0.163
$\leq$ 100 v. > 100	-0.3293	0.1804	0.076
Diagnosis (required v. not required)	-0.1621	0.1800	0.374
Study quality (high v. medium)	-0.3371	0.2174	0.130
Fidelity measure	-0.3381	0.2288	0.153
Treatment length	0.0143	0.0161	0.380
Control group (no TR/waitlist v. TAU/other therapy)	-0.5160	0.1600	**0.003****

Table 4 – Univariate analysis results of total studies
(* $p < 0.05$, ** $p < 0.01$)

3.3.2 Multivariate meta-regression analysis

In the analysis with all of the variables from 22 eligible studies, only the fidelity measure was significant (β = -0.86, s.e.(β) = 0.30, p = 0.017; analysed with dichotomous versions of age, participation rate, and sample size variables). Fidelity measure remained significant ($p < 0.05$) in each of the combinations that included one version of the redundant variables except with continuous versions of sample size and attrition rate (p = 0.077), which still approached significance. Studies that did not clearly state the use of a fidelity measure were entered as missing data in the initial analysis. Even when these cases were assumed

as absence of a fidelity measure, the variable remained significant in 32 studies (β = -0.52, s.e.(β) = 0.23, p = 0.033).

3.3.3 Exploratory subgroup analysis

To check whether findings were similar within the subsets of studies with same control group type, additional meta-regression analyses were conducted: sample size (both continuous and dichotomous versions) and participation rate (both versions) were significant in the univariate analysis of studies with no treatment/waitlist controls ($p < 0.05$). Publication year approached significance (p = 0.074). When these variables were examined together in the multivariate analysis, participation rate (β = -0.58, s.e.(β) = 0.20, p = 0.011) and sample size (Continuous: β = -0.004, s.e.(β) = 0.002, p = 0.050; dichotomous: β = -0.50, s.e.(β) = 0.18, P = 0.015) remained significant.

None of the variables in studies with TAU/other therapy controls was significant in the univariate meta-regression analysis and thus, no multivariate analysis was conducted for this subgroup.

3.4 Difficulties of conducting RCTs in prisons (qualitative analysis)

One of the most commonly identified difficulties in implementing RCTs of psychological therapy in prison was the difficulty in long term follow-up [34,66,82,88,96,99] due to high rate of release,[88] rapid turnover of prisoners,[83] and short duration of stay.[85] Short treatment period was a similar limitation.[66] One author commented that combination of having a small sample, volunteer participants, and time constraints prevented the collection of post-test follow-up measures.[96]

Institutional constraints on scheduling, which prevented the implementation of certain study measures, were also frequently mentioned.[33,96] For example, collection of biological

markers of stress were prohibited in one study.[33] Ford et al. (2012) details the extent of institutional constraints in their study as the following:

> Department of Correction policy required that participants receive both interventions without any delay, thus preventing follow-up assessments to ascertain longer-term outcomes. Some benefits of the skills-based TARGET intervention also may not be evident until recipients have an extended period to use the skills (Ford et al., 2011), but that could not be tested in this study. In addition, institutional policies prohibiting the taping of therapy sessions prevented the independent collection of fidelity ratings and reliability check. Treatment recipients may have received a suboptimal "dose" of therapy because they attended on average only two-thirds of sessions. Most nonattendance was due to unavoidable scheduling problems that are expectable in prisons.[76]

Similarly, Cole et al. (2007) commented that institutional constraints limited the recruiting time for participants by slowing down institutional review board (IRB) approval process.

Moreover, balancing clinical needs of the studies with the scientific needs were also one of the named difficulties. For example, one author commented that a placebo control would improve the study's scientific design but would not be approved by the review boards.[82] Having the same clinicians for both conditions due to limited resources was also described as a limitation.[79] In another study one therapist carried out the intervention for all participants in the treatment group, which is another drawback as it leaves the possibility that the intervention response could be therapist-specific.[95]

Controlling for social environment of the prison[75,88,95] was also reported to be difficult. Biggam and Power (2002) stated that "the role that subtle changes to the regime or the ethos in the institution played in the prisoners' welfare cannot be ascertained and evaluated."[95] Another author commented that potential contamination of treatment and

control conditions may have occurred due to the closed communal setting of the prison wing.[79] Messina et al. (2010) commented that programming hours in prison are often limited or interrupted by security issues such as "lock downs," which needs to be overcome by training peer mentors to enable prisoners to continue their treatment exercises when restricted to their housing units.

Having a small participant sample was also a highlighted.[34,66,68,74-76,78,96] Reasons for obtaining a small sample included difficulty in participant recruitment,[96,100] and in one study with a juvenile sample parental non-consent was an issue as well.[86] Similarly, high attrition rate[68,78,83,84,87,98] partly due to schedule changes[96,98] and inmate infractions that restricted program enrolment[98] were often mentioned as difficulties. Sometimes control group participants were more likely to drop out of the study.[78] One author commented that the disruptiveness of the facility may have increased the attrition rate.[84]

Reliance on self-report measures was often identified as drawbacks in studies of psychological therapies conducted in prisons.[62,65,74,80,87,98] Bradley & Follingard (2003) stated that there may be a possibility that "inmates wanted to 'fake good' in order to be considered parole or placement evaluation," even though the authors believed it was not very likely. Not blinding the participants to treatment condition[65] was a further constraint on the study validity.

Authors also cited difficulty in ensuring continuity of care as a limitation.[64,86] Non-attendance after release[86] was one cause, while another study reported that clinical staff of the prison could not produce the conditions of safety and confidentiality that they had with facilitators provided by the study, which may limit generalisability of the study findings.[64]

3.5 Publication bias

Funnel plot analysis with the Egger's test did not show evidence of publication bias with the total set of studies (t =1.83, s.e.(t) = 0.90, p = 0.075). However, when tests were performed again with the subgroup with no treatment/waitlist controls, results suggested strong evidence of publication bias (t =3.16, s.e.(t) = 1.00, p = 0.005) (Appendix J). A subset of studies with TAU/other therapy controls did not show evidence of publication bias (t = -0.96, s.e.(t) = 1.30, p = 0.36).

As there was no evidence of funnel plot asymmetry in the total set of studies, the trim and fill method did not produce a different pooled effect estimate for the entire set. Even when this analysis was done on the subgroup with no treatment/waitlist controls, it produced the same effect estimate compared to the actual pooled effect size of the 21 studies, with only a 0.001 unit of decrease in the upper limit of the 95% confidence interval.

Chapter 4: Discussion

4.1 Overview

The aim of this thesis was to present a systematic review and meta-analysis of mental health outcomes of RCTs of psychological therapies in prisons, based on 37 studies involving 2,761 prisoners. This final chapter discusses the findings and assesses the strengths and limitations of this review. In addition, quality appraisals of included studies are reflected upon. The implications of the findings and directions for future research are also considered.

4.2 Principal findings

The pooled effect size of the included RCTs was 0.50 (95% CI: 0.34–0.66) with moderate to high levels of heterogeneity (I^2 = 73%, 95% CI: 62%–80%). This effect size is similar to that of psychotherapy used in the community setting for acute mental disorders, according to an umbrella review that primarily included CBTs, disorder-specific therapies, and few psychodynamic approaches (ES = 0.58, 95% CI: 0.40–0.76).[122] The effect size reported in this review could be described as medium, whereas a threshold of 0.8 or more would be large.[121] Moreover, as I^2 of 75% indicate high levels of heterogeneity[123], the individual effect sizes of each included studies would show high percentage of variation owning to true heterogeneity. Such level of heterogeneity was expected due to the clinical differences previously mentioned (section 2.4.2), and therefore, the pooled effect size could be viewed as a comparison point for each subgroup rather than "a true effect estimate" that is expected in a fixed effect meta-analysis.[69]

In the subgroup analyses, the effect size of the studies with no treatment/waitlist controls was 0.74 (95% CI: 0.51–0.97; I^2 = 74%, 95% CI: 60%–83%) and that of studies with TAU or other psychological therapy controls was 0.21 (95% CI: 0.01–0.42; I^2 = 63%, 95% CI: 36%–78%). This difference was also found in community settings; mean effect size for the subgroup with waitlist controls was 0.86 and that of the subgroup with active treatment control was 0.46 (U = 5.00, p = 0.003; from 8 meta-analyses).[122] However, one important difference is that therapies in the current review were conducted for most participants who did not require diagnosis, while the meta-analysis on the general population covered acute treatments.

Both subgroups in these analyses showed moderate to high levels of heterogeneity, which is expected due to the differences in study characteristics other than the control group type. The wider confidence interval of the studies TAU/other therapy controls is likely due to the lower number of studies (n = 16) than those with no treatment/waitlist controls (n = 21). Slightly higher levels of heterogeneity in the no treatment/waitlist control subgroup can be due to the greater number of included trials and more variations in study conditions across them.

Moreover, 3-month follow-up and 6-month follow-up results show that therapy trials in prison need to improve on the tracking of follow-up outcomes, as only a small portion of studies included in this review (6 out of 37) recorded follow-up measures. Furthermore, the result from this review suggests that psychological therapies in prisons need improvement in maintaining the initial treatment gains, as no significant differences were found between the treatment groups and the control groups at the follow-up. There is a possibility that a significant effect was not detected in this analysis due to limited power,

but it can be suspected that any treatment effect that actually exists is not large, as it was not picked up in the current set of analyses.

When studies were grouped by treatment type, there were no significant differences among the various categories of therapies. This suggests that the four categories of treatment covered in this review did not significantly differ in their efficacy in prisons. However, there was a consistent trend in the results that could suggest that trauma-based therapies had a smaller pooled effect size (ES = 0.27, 95% CI: -0.12–0.65) than the rest while mindfulness-based therapies tended to have a larger effect size (ES = 0.83, 95% CI: 0.38–1.28). Therefore, while the difference between treatment types is not statistically significant, a visual analysis of the results suggests that trauma-based therapies would benefit from more research to improve the effectiveness of treatment. The need for such improvement is supported by the fact that trauma symptom outcomes also tended to have smaller effect sizes (ES = 0.35, 95% CI: 0.14–0.56) than rest of the specific psychological outcomes. Moreover, improvement in these therapies would help target the current unmet needs of prisoners, given that prisoners are a population with high levels of PTSD prevalence and trauma history.[124-127]

Results also suggest that overall psychopathology outcomes (ES = 0.32, 95% CI: 0.05–0.59) had a smaller effect size than other types of outcomes (depression: ES = 0.60, 95% CI: 0.38–0.83, I^2 = 71%, 95% CI: 54%–81%; anxiety: ES = 0.57, 95% CI: 0.31–0.84, I^2 = 82%, 95% CI: 74%–88%), which may indicate the current difficulty with using these therapies to substantially improve the status of overall psychopathology compared to improving just depression or anxiety symptoms.

As mentioned, some therapies in the others category were developed by combining existing types such as CBT and mindfulness-based therapy. Combining two techniques is

naturally expected to benefit from the strengths of both techniques, unless they are incompatible. The two therapies in this review that combined both CBT and mindfulness concepts, however, had lower effect sizes than the pooled effect sizes of either CBT-based or mindfulness-based categories.[68,77] Therefore, these studies are unlikely to have elevated the pooled effect size of the "others" category by not being placed into the CBT or mindfulness-based categories. Further research could explore the efficacy of these combined techniques in comparison with treatments that employ only one of the techniques.

Similar to the cases in the main analysis, moderate to high levels of heterogeneity in the secondary analyses suggest that pooled effect sizes of each category of studies has some limitation in representing the treatment effect of individual studies. There were some cases in which the estimated level of heterogeneity was low (i.e. trauma outcome and follow-up outcomes). However, since their 95% CI of I^2 span as widely as 0% to 82%, there is limited reliability in the estimated level of true heterogeneity. Such wide confidence intervals were produced for categories including less than ten studies, and consequently estimated pooled effect sizes for these categories should be interpreted with caution.

In addition, the finding that studies did not significantly differ in effect size by participation type could support that group therapies may be more efficient in the prison setting as it requires less treatment resources (i.e. treatment providers)—except for certain treatments for acute disorders that may require more individual attention and confidentiality during the therapy process. Studies that combine both group and individual elements could be more explored in future research, as the number of studies found in this review was more limited than those of other participation types.

4.3 Meta-regression analysis results

The fact that the presence and the strength of a treatment fidelity measure were associated with lower effect size in the multivariate meta-regression is surprising at first glance. Fidelity measures are implemented to ensure consistent quality and delivery of treatment following the original program model.[128] Therefore, studies that use sufficient fidelity measures are expected to be better able to consistently deliver treatment of an adequate quality to their participants. The results of this analysis do not necessarily suggest a causal relationship between the implementation of fidelity measures and effect size. A possible explanation for this finding is that studies that implemented an adequate fidelity measure may have been studies that had more stringent study conditions than those that did not. Thus, they may have tended to report less inflated effect sizes. In other words, studies that did not implement any fidelity measures or a measure of sufficient quality may be reporting inflated effect sizes produced by sources of bias such as lack of blinding of researchers and participants. To check whether this hypothesis has any identifiable support from the current data, studies with adequate fidelity measures[64,76,78,85-87,95-97,100] were examined in more detail: there was indeed some overlap with the studies of higher quality rating.[75-77,82,83,86,100] There was also certain overlap with studies that had the TAU/other therapy controls.[32,62,65-67,74-79,83,85,86,88,97] It is notable that except one study,[97] studies of older publication dates (publication year < 2000) did not have an adequate fidelity measure or did not state whether they did.[66,74,81,88]

Variables that were significant in the univariate meta-regression analysis (e.g. participation rate and control group type) also suggest that larger effect sizes may be associated with factors that could inflate the effect size such as attrition (outcomes from remaining participants are likely to be higher than those from participants who dropped out) and a control group that does not receive any active treatment.

Furthermore, it is important to note that meta-regression results were different for each subset of studies divided by control group type. In the no treatment or waitlist control studies, participation rate and sample size remained significant in the multivariate regression while studies with TAU or other control studies did not show any evidence of such variables' linear relationship with the effect size. This result may have been found because there was lower estimated level of heterogeneity in this subgroup (I^2 = 63%, 95% CI: 36%–78%) than that of studies with a no treatment/waitlist controls (I^2 = 74%, 95% CI: 60%–83%). In other words, there may have been less differences between these studies because there were fewer variations in study design and consequently, in efficacy among studies with a stricter controls. This finding also supports the claim that the association of effect size with variables such as fidelity measure, sample size, and participation rate are not found when conditions of the treatment trial are stricter and follow higher standards of preventive measures to minimize bias in the study.

Other possible sources of heterogeneity that were not accounted for in the meta-regression analysis include differences in baseline outcome measures: in some studies, there were higher baselines in the control group while in others there were higher baselines in treatment group.

4.4 Treatment length

It is notable that most of the studies included in this review were short-term treatments. As the average treatment length was 10 weeks, which is about two months, it can be seen that many studies selected a short-term treatment to accommodate the prison setting and the research conditions. Providing short-term psychological therapies can be efficient, as the length of treatment does not seem to influence the treatment effect size according to the meta-regression results. Regardless of treatment length, however, efforts to improve

maintenance of treatment gains are necessary since outcomes from this review suggest that the treatment effect is not significant at follow-up.

4.5 Difficulties of conducting RCTs in prisons

As a considerable portion of the difficulties seems to arise from the structure and regulations of prisons, much of these obstacles would not be overcome simply with the best efforts on the part of researchers who conduct these treatment trials. Getting in contact with the relevant staff and department of the prisons well in advance of starting the treatment would be helpful in reducing factors that disrupt the treatment practice or data collection. There were also many shortcomings in the trials that could be improved if more resources were available to the researchers: small sample was a frequent limitation across the included studies, but other studies showed that recruitment of a substantial sample (>150) in prisons was possible by using a multi-site format of study design.[32,33,64,83] In multi-site trials, adherence to the protocol of the study in all sites must be thoroughly checked to ensure that the results are comparable in different sites. In addition, analysis on the comparability of the results from multiple sites should be performed before reporting combined outcomes. One of the suggested solutions for the problem of limited resources available for psychological therapies in prisons was utilizing more behavioural-oriented interventions.[33] For example, yoga, art therapy, and music therapy could be more cost-effective in the prison setting,[33] especially because it could be more accessible and less stigmatizing among prisoners than other psychosocial treatments.[129-131] The validity of this opinion should be further explored in research on behavioural therapies in prisons.

4.6 Publication bias

Evidence for publication bias was found in the subset of studies with no treatment/waitlist controls but not in rest of the studies or when all of the included studies were examined as a whole. This result suggests that publication bias or reporting bias may be found in the

subset of studies with no treatment/waitlist controls. It is possible that selective outcome reporting or selective analysis reporting of individual studies could have caused the asymmetry in the funnel plot of the subgroup. There are also other possible explanations. These include: poor methodological quality leading to "spuriously inflated effects in smaller studies," true heterogeneity, sampling variation, or chance.[92]

4.7 Strengths & limitations

4.7.1 Overview of strengths

This review is unique in that it includes the first meta-analyses of outcomes of psychological therapies in prisoners (of both genders) excluding pharmacological therapy to my knowledge. As previously mentioned, this review was able to include the most number of RCTs of psychological therapies in prison compared to former reviews that examined mental health gain in prisoners. Furthermore, while the individual findings of studies with small samples are limited by statistical power when examined separately, the meta-analysis conducted in this review was able to combine these small studies, including pilot studies, to produce results that have much greater statistical power. Therefore, this systematic review was able to help increase the validity of the conclusions that can be drawn from the results of the studies.

4.7.2 Study quality & characteristics

Adopting the methodology of a systematic review had advantages and also some disadvantages. A strength of this review is that it used a transparent and reproducible search strategy and has been registered on the international database of prospectively registered systematic reviews in health and social care.[60] However, the quality of this systematic review inevitable depends, at least partially, on the quality of the included studies.

Many of the included studies lack the double-blinding procedure—but the double-blind practice would be difficult given the conditions of a psychological treatment study,[122] especially in prisons. Therefore when it was single-blinded (blind researchers), the study was rated as 1 (well-addressed) on the item of double-blinding in the quality checklist. The studies meeting this criterion were only 7 out of 37 studies. Other four studies commented on the lack of blinding or their justification for not adopting the double-blind procedure, but rest of the 26 studies did not mention the blinding procedure at all. Therefore, this is clearly a limitation of many of the studies included in this review, as lack of blinding can favourably bias the treatment condition. Imperfect blinding has been a commonly identified issue in other meta-analyses of psychotherapy studies as well.[122]

Moreover, 8 out of 37 studies did not employ the intention-to-treat (ITT) analysis. This could also have favourably biased the treatment group since it is possible that participants that dropped out of the treatment group may have reported lower treatment outcome if they were able to provide the data.

Authors of individual studies frequently mentioned self-report measures as possible limitations of their studies.[62,65,74,80,87,98] Even with the potential restrictions of self-reports, these types of measures seemed to be appropriate in general for the psychological therapies in this review; clinical interviews that only check for presence or absence of a formal diagnosis could be more limited in representing the treatment effect, especially since many of the participants of the studies were not required to have a diagnosis at the enrolment of treatment. Moreover, an assessment of psychological well-being is inherently dependent on the communication by the individuals being assessed, unlike physical measures that could be collected though strictly objective measures. In addition, it is important that the individuals themselves experience and benefit from the treatment effects—therefore, self-

report measures were deemed to be largely suitable for measuring the treatment effect of psychological therapies in this review.

Moreover, another possible limitation is that no studies outside the UK, the United States, China, Iran, Norway, India and Spain were included in this review. Over half of the studies were from the U.S. (26 studies), which may also limit the generalisability of these findings. Therefore, the need for research in other countries is highlighted.

Despite the stated limitations in study qualities, all of studies included were rated above medium quality according to the NICE guidelines quality checklist. Treatment fidelity was not included in the quality checklist, but it was reported separately and analysed in the meta-regression analysis.

Although this review set randomisation as the inclusion criteria, one included study had participants that were randomised and also participants who chose the treatment type by preference.[64] This study compared Seeking Safety intervention and M-TREM (Male-Trauma Recovery Empowerment Model) and a waitlist control group. The justification for the inclusion of this study is that the authors controlled for between group differences in preference assignment v. random assignment by entering level-2 coefficients in their analysis. They have also stated that preference assignment showed no significant effect on the outcomes. Moreover, the authors state that this study actually may have more strength in their outcome by including preference assignment, so it may better reflect the settings in the community-based applications as opposed to the limited generalisability of the clinical trials.

In fact, including only RCTs in this review could be a drawback to applicability of the findings to psychological treatments regularly provided in prisons: unlike the conditions in RCTs, participants may be mandated to treatment in cases of substance abuse (for

committers of substance-related crimes), or in other cases, would be able to enrol in the type of treatment they prefer without being assigned to a specific type. Since active engagement in treatment is critical in increasing the effectiveness of psychological therapies,[132,133] in studies that compared two different types of treatments, random assignment may have lowered the treatment effect by decreasing participants' motivation to engage in treatment if they were not able to enrol in the program they wanted.

Moreover, if a participant requires treatment for another disorder, it would be difficult to pause any other treatment of diagnosed patients even if they are placed in the control group. Such problems can arise especially in prisons, given the high rate of co-morbidity of psychological disorders in prisons. To solve for such ethical issues and minimise confounding factors of additional treatment, some studies provided both the control group and treatment group with valid treatments. However, such results may underreport treatment effects for such studies compared treatment with no treatment or waitlist controls. While RCTs included in this review mostly did not require diagnosed patients, treatment for such patients in prison may be better assessed with clinical trials that examine an individual treatment on its own without a control group.

Despite such possible limitations from including only RCTs, I believe the current review still provides important findings that can guide the practice of psychological treatments for undiagnosed prisoners with the benefits of randomisation and control of study conditions.

4.7.3 Methodology of research – search terms

The search terms included in this review specifically targeted for certain types of psychological therapy (i.e. CBT) and certain types of disorders (i.e. depression with "depress*" and schizophrenia with "schizo*") while rest of the therapy types and disorders were aimed to be covered by wildcard search terms such as "mental*," "psych*,"

"disorder*," "group therap*," "intervention*," and "treatment*." While I believe most psychological disorders would have been able to be identified with these search terms, they may have missed some RCTs targeted at undiagnosed psychological difficulties such as anxiety (and not anxiety disorder). The most number of included studies had depression outcomes compared to other specific psychological outcomes (i.e. anxiety, trauma, and overall psychopathology). Specifically targeting for depression may have contributed to this result, although the fact that depression is one of the most commonly treated psychological difficulties would have contributed as well. Additional targeted searches were also done with the search terms with "therap*" and "psychotherapy" to identify any therapies not covered with the initial search terms. Therefore, it would be useful to include these terms in the main search in future investigations.

4.7.4 Effect size calculation

In this review, post-treatment scores were used to calculate the effect size. According to the *Cochrane Handbook*, the guide for systematic review of health interventions, using analysis of covariance (ANCOVA) by looking at post-treatment outcomes using baseline scores as a covariate would yield the most precise and least biased estimate of the treatment effects. However, this analysis was not available from most of the studies included in this review. One study did provide a Cohen's d adjusted with the baseline scores as a covariate,[75] but the effect size calculated from post-test scores was used in the analyses of this review in order to apply uniform methods to all studies. Studies that had exceptionally higher effect sizes than rest of the studies (ES > 0.90) were individually examined: one study reported mean change scores[63] (another did in their paper but calculation was done with additional data on post-test scores provided by the author)[34] and four studies reported lower baselines scores (i.e. lower severity of mental health problems) of the treatment group,[80,81,95,96] which may have inflated the effect sizes of these studies.

Two studies that did not fall into either category were a meditation study and art therapy study.[34,62] These studies with higher effect sizes were distributed across the different modes of therapy and participation types, and therefore, unlikely to have contributed to falsely attributing a higher pooled effect size in any specific type of treatments.

4.7.5 Variation in outcome types

For one study the scale for measuring sleeping difficulties were used even though it was a scale created by the authors and was not validated outside of the study.[62] Although this might be a limitation, the study was still included as it was using a standardised measure and not a narrative interview or other unquantifiable measures. Similarly, in the secondary analysis, the number of post-treatment infirmary visits was used for one study in the analysis of physical health outcome measures.[67] While this measurement is limited compared to other types of standardised measurements, it was only included in the exploratory analysis, which did not yield a significant result.

It is important to note that this review incorporated various standardised psychological measures, instead of unifying the outcome measure to one specific type. Although some may say that this could be a limitation, this method was chosen for the main analysis to include the most number of studies and because the standardisation of mean differences allows for valid comparisons of effects originally measured on different scales.[28] The primary outcome measure was chosen for each study to obtain the most overlap among the included studies, so the possible bias created by handpicking outcome measures that produced the largest treatment effect for each study was minimized. Furthermore, additional analyses looking at specific types of psychological outcome were conducted with subsets of studies that provided the outcome. Nevertheless, measurements varied even in the additional analyses, so the heterogeneity possibly arising from different validity and reliability of each measurement could not be fully accounted for.

4.7.6 Heterogeneity in studies

There was a considerable level of heterogeneity in the total set of included studies (I^2 = 73%, 95% CI: 62%–80%), which %), which mean that our overall effect size should be interpreted with caution. To deal with this limitation, we conducted subgroup analyses and meta-regression to examine sources of heterogeneity. However, there could have been other sources of heterogeneity that were not reported and were not tested. High levels of heterogeneity are not unusual for meta-analyses of RCTs[11,17,134] and partly reflect the diverse populations being studied.[73]

4.8 Clinical significance of findings

The difference between treatments for psychological problems in prison to those in the community is that not all types of treatment are accessible in each prison and that those in prison are inevitably subject to stressors in the environment that cannot be eliminated (such as deprivation of liberty, being away from family, restricted work opportunities). Therefore, the findings from RCTs of psychological treatments conducted in prisons could be important for practitioners and policy makers to guide the commissioning of future services and managing existing ones.

The main findings of the current review provide some support for the use of psychological therapies in prisons, but the evidence is not definitive for what type of psychological therapy is most effective for all mental health conditions and problems. In addition, the results of this synthesis suggest that treatment lengths of longer than 8 weeks were not associated with improved outcomes.

An implication of the lack of treatment specificity is that group therapy could be offered in a first instance, as it is more cost-effective. At the same time, research needs to clarify if

combined individual and group therapies are better as the individual studies included in this review that used combination approaches were too variable to draw firm conclusions.

There was some tentative evidence that therapies aimed at improving traumatic symptoms were not as effective as those treatments that targeted depression or anxiety, which suggests that more research is needed into the trauma-based treatments in prisoners. In addition, some preliminary findings suggested that mindfulness and CBT-based therapies have evidence for treating depression or anxiety in prisons. It is likely that "topping-up" or additional treatment sessions after end of a treatment program may be necessary in order to maintain the treatment effect based on the results finding little evidence of treatment effects lasting after therapies have ended.

4.9 Conclusion

While psychological therapies in prisons tend to show medium treatment effect compared to no treatment or waitlist group, the effect is lower when the controls are given treatment-as-usual or another form of psychological therapy. Whether this level of evidence is sufficiently strong for the introduction of such therapies in prison requires further discussion and better quality studies going forward. Results also suggest that studies should adopt an adequate fidelity measure, and caution is necessary in interpreting studies with small sample sizes (≤ 100) and high attrition rates ($\geq 20\%$). Findings also show that improvement in retaining the initial treatment gains may be needed.

References

1. Walmsley R. World Population List: International Centre for Prison Studies, 2013.
2. Fazel S, Seewald K. Severe mental illness in 33 588 prisoners worldwide: systematic review and meta-regression analysis; 2012.
3. Bromet E, Andrade L, Hwang I, et al. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Medicine* 2011; **9**(1): 90.
4. Fazel S, Baillargeon J. The health of prisoners. *The Lancet* 2011; **377**(9769): 956-65.
5. Butler T, Andrews G, Allnutt S, Sakashita C, Smith NE, Basson J. Mental disorders in Australian prisoners: a comparison with a community sample. *Australian and New Zealand Journal of Psychiatry* 2006; **40**(3): 272-6.
6. Butler T, Allnutt S. Mental illness among New South Wales prisoners: Corrections Health Service Sydney, NSW; 2003.
7. Blaauw E, Roesch R, Kerkhof A. Mental Disorders in European Prison Systems: Arrangements for Mentally Disordered Prisoners in the Prison Systems of 13 European Countries. *International Journal of Law and Psychiatry* 2000; **23**(5–6): 649-63.
8. Brink JH, Doherty D, Boer A. Mental disorder in federal offenders: A Canadian prevalence study. *International Journal of Law and Psychiatry* 2001; **24**(4–5): 339-56.
9. Fazel S, Benning, Ram,, Danesh J. Suicides in male prisoners in England and Wales, 1978–2003. *The Lancet* 2005; **336**(9493): 8-14, 1301-2.
10. Fazel S, Benning R. Suicides in female prisoners in England and Wales, 1978–2004. *The British Journal of Psychiatry* 2009; **194**(2): 183-4.
11. Fazel S, Benning, Ram, & Danesh, John. Suicides in male prisoners in England and Wales, 1978–2003. *The Lancet* 2005; **336**(9493): 8-14, 1301-2.
12. Hawton K, Linsell L, Adeniji T, Sariaslan A, Fazel S. Self-harm in prisons in England and Wales: an epidemiological study of prevalence, risk factors, clustering, and subsequent suicide. 2014; **383**(9923): 1147-54.
13. Fazel S, Cartwright J, Norman-Nott A, Hawton K. Suicide in prisoners: a systematic review of risk factors. 2008; **69**(11): 1721-31.
14. Rivlin A, Hawton K, Marzano L, Fazel S. Psychiatric disorders in male prisoners who made near-lethal suicide attempts: case–control study. *The British Journal of Psychiatry* 2010; **197**(4): 313-9.
15. Baillargeon J, Binswanger IA, Penn JV, Williams BA, Murray OJ. Psychiatric Disorders and Repeat Incarcerations: The Revolving Prison Door. *American Journal of Psychiatry* 2009; **166**(1): 103-9.
16. Coid JW, Ullrich S, Keers R, et al. Gang Membership, Violence, and Psychiatric Morbidity. *American Journal of Psychiatry* 2013; **170**(9): 985-93.
17. Bartlett A, Jhanji E, White S, Anne Harty M, Scammell J, Allen S. Interventions with women offenders: a systematic review and meta-analysis of mental health gain. *The Journal of Forensic Psychiatry & Psychology* 2014; **26**(2): 133-65.
18. Brooker C, Sirdifield C, Gojkovic D. Mental health services and prisoners: A updated review. Lincoln, UK: CCAWI University of Lincoln; 2007.
19. Fontanarosa J, Uhl S, Oyesanmi O, Schoelles KM. Interventions for Adult Offenders With Serious Mental Illness: Agency for Healthcare Research and Quality (US), Rockville (MD); 2013.
20. Heckman CJ, Cropsey KL, Olds-Davis T. Posttraumatic stress disorder treatment in correctional settings: A brief review of the empirical literature and suggestions for future research. *Psychotherapy: Theory, Research, Practice, Training* 2007; **44**(1): 46-53.

21. Himelstein S. Meditation Research: The State of the Art in Correctional Settings. *International Journal of Offender Therapy and Comparative Criminology* 2011; **55**(4): 646-61.
22. Kouyoumdjian FG, McIsaac KE, Liauw J, et al. A Systematic Review of Randomized Controlled Trials of Interventions to Improve the Health of Persons During Imprisonment and in the Year After Release. *American Journal of Public Health* 2015; **105**(4): e13-e33.
23. Leigh-Hunt N, Perry A. A Systematic Review of Interventions for Anxiety, Depression, and PTSD in Adult Offenders. *International Journal of Offender Therapy and Comparative Criminology* 2014.
24. Morgan RD, Flora DB. Group psychotherapy with incarcerated offenders: A research synthesis. *Group Dynamics: Theory, Research, and Practice* 2002; **6**(3): 203-18.
25. Ross J, Quayle E, Newman E, Tansey L. The impact of psychological therapies on violent behaviour in clinical and forensic settings: A systematic review. *Aggression and Violent Behavior* 2013; **18**(6): 761-73.
26. Hockenhull JC, Whittington R, Leitner M, et al. A systematic review of prevention and intervention strategies for populations at high risk of engaging in violent behaviour: update 2002-8. *Health Technol Assess* 2012; **16**(3): 1-152.
27. Shonin E, Van Gordon W, Slade K, Griffiths MD. Mindfulness and other Buddhist-derived interventions in correctional settings: A systematic review. *Aggression and Violent Behavior* 2013; **18**(3): 365-72.
28. Higgins JPT GS. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 ed: The Cochrane Collaboration. ; 2011.
29. Borenstein MH LVH, Julian P.T.; Rothstein, Hannah R. Introduction to meta-analysis. Croydon: John Wiley & sons, Ltd.
30. Shapiro S. Meta-analysis/Shmeta-analysis. *American Journal of Epidemiology* 1994; **140**(9): 771-8.
31. Egger MS, George Davey; Altman Douglas G. Systematic Reviews in Healthcare. London: BMJ Publishing Group; 2007.
32. Chen X-J, Hannibal N, Gold C. Randomized Trial of Group Music Therapy With Chinese Prisoners: Impact on Anxiety, Depression, and Self-Esteem. *International Journal of Offender Therapy and Comparative Criminology* 2015.
33. Bilderbeck AC, Farias M, Brazil IA, Jakobowitz S, Wikholm C. Participation in a 10-week course of yoga improves behavioural control and decreases psychological distress in a prison population. *Journal of Psychiatric Research* 2013; **47**(10): 1438-45.
34. Gussak D. The Effectiveness of Art Therapy in Reducing Depression in Prison Populations. *International Journal of Offender Therapy and Comparative Criminology* 2007; **51**(4): 444-60.
35. Law FM, Guo GJ. Hope and recovery from substance abuse for female drug offenders in Taiwan. *International Journal of Offender Therapy and Comparative Criminology* 2012; **56**(8): 1258-82.
36. Sacks JY, McKendrick K, Hamilton Z. A Randomized Clinical Trial of a Therapeutic Community Treatment for Female Inmates: Outcomes at 6 and 12 Months After Prison Release. *Journal of Addictive Diseases* 2012; **31**(3): 258-69.
37. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery* 2010; **8**(5): 336-41.
38. Cosden M, Ellens JK, Schnell JL, Yamini-Diouf Y, Wolfe MM. Evaluation of a mental health treatment court with assertive community treatment. *Behavioral sciences & the law* 2003; **21**(4): 415-27.

39. Bichescu D, Neuner F, Schauer M, Elbert T. Narrative exposure therapy for political imprisonment-related chronic posttraumatic stress disorder and depression. *Behaviour Research and Therapy* 2007; **45**(9): 2212-20.
40. Nancee Blum MSW, Don St. John MA, P.A.C., Bruce Pfohl MD, et al. Systems Training for Emotional Predictability and Problem Solving (STEPPS) for Outpatients With Borderline Personality Disorder: A Randomized Controlled Trial and 1-Year Follow-Up. *American Journal of Psychiatry* 2008; **165**(4): 468-78.
41. Rocha NBF, Marques AB, Fortuna RB, Antunes A, Hoaken PNS. Effectiveness of cognitive remediation for female inmates: a pilot study. *The Journal of Forensic Psychiatry & Psychology* 2014; **25**(2): 224-37.
42. Shelton D, Sampl S, Kesten KL, Zhang W, Trestman RL. Treatment of impulsive aggression in correctional settings. *Behavioral sciences & the law* 2009; **27**(5): 787-800.
43. Black DW, Blum N, Eichinger L, McCormick B, Allen J, Sieleni B. STEPPS: Systems Training for Emotional Predictability and Problem Solving in women offenders with borderline personality disorder in prison—A pilot study. *CNS spectrums* 2008; **13**(10): 881-6.
44. Simpson TL, Kaysen D, Bowen S, et al. PTSD symptoms, substance use, and vipassana meditation among incarcerated individuals. *Journal of Traumatic Stress* 2007; **20**(3): 239-49.
45. Nee C, Farman S. Dialectical behaviour therapy as a treatment for borderline personality disorder in prisons: Three illustrative case studies. *The Journal of Forensic Psychiatry & Psychology* 2007; **18**(2): 160-80.
46. Echeburúa E, Fernández-Montalvo J, Amor PJ. Psychological Treatment of Men Convicted of Gender Violence: A Pilot Study in Spanish Prisons. *International Journal of Offender Therapy and Comparative Criminology* 2006; **50**(1): 57-70.
47. Doyle M, Khanna T, Lennox C, et al. The effectiveness of an enhanced thinking skills programme in offenders with antisocial personality traits. *The Journal of Forensic Psychiatry & Psychology* 2012; **24**(1): 1-15.
48. Dickens JE. Evaluating the effectiveness of a manualized treatment for inmates with dual diagnoses [Ph.D.]. Ann Arbor: University of Nevada, Las Vegas; 2007.
49. Orme-Johnson DW, Moore RM. First Prison Study Using the Transcendental Meditation Program. *Journal of Offender Rehabilitation* 2003; **36**(1-4): 89-95.
50. Chandler D, Spicer G. Integrated Treatment for Jail Recidivists with Co-occurring Psychiatric and Substance Use Disorders. *Community Ment Health J* 2006; **42**(4): 405-25.
51. Bowen S, Witkiewitz K, Dillworth TM, et al. Mindfulness meditation and substance use in an incarcerated population. *Psychology of Addictive Behaviors* 2006; **20**(3): 343-7.
52. Abracen J, Looman J, Langton CM. Treatment of Sexual Offenders With Psychopathic Traits: Recent Research Developments and Clinical Implications. *Trauma, Violence, & Abuse* 2008; **9**(3): 144-66.
53. Müller-Isberner R, Hodgins S. Violence, crime and mentally disordered offenders. Chichester, UK: Wiley; 2000.
54. Salekin RT, Worley C, Grimes RD. Treatment of psychopathy: A review and brief introduction to the mental model mpproach for psychopathy. *Behavioral sciences & the law* 2010; **28**(2): 235.
55. Harris GT, Rice ME. Treatment of psychopathy. *Handbook of psychopathy* 2006: 555-72.
56. Salekin RT. Psychopathy and therapeutic pessimism: Clinical lore or clinical reality? *Clinical Psychology Review* 2002; **22**(1): 79-112.

57. Kubiak SP, Fedock G, Tillander E, Kim WJ, Bybee D. Assessing the feasibility and fidelity of an intervention for women with violent offenses. *Evaluation and Program Planning* 2014; **42**(0): 1-10.
58. Lee K-H, Bowen S, An-Fu B. Psychosocial outcomes of mindfulness-based relapse prevention in incarcerated substance abusers in Taiwan: A preliminary study. *Journal of Substance Use* 2011; **16**(6): 476-83.
59. (UK) NCCfMH. Borderline Personality Disorder: Treatment and Management. Leicester UK: The British Psychological Society & The Royal College of Psychiatrists.; 2009.
60. PROSPERO - International prospective register of systematic reviews. http://www.crd.york.ac.uk/PROSPERO/.
61. Wilson DB. Practical meta-analysis effect size calculator. 2001. http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD1.php (accessed 2014-01-23.
62. Sumter MT, Monk-Turner E, Turner C. The Benefits of Meditation Practice in the Correctional Setting. *Journal of Correctional Health Care* 2009; **15**(1): 47-57.
63. Khodayarifard M, Shokoohi-Yekta M, Hamot GE. Effects of Individual and Group Cognitive-Behavioral Therapy for Male Prisoners in Iran. *International Journal of Offender Therapy and Comparative Criminology* 2010; **54**(5): 743-55.
64. Wolff N, Huening J, Shi J, Frueh BC, Hoover DR, McHugo G. Implementation and effectiveness of integrated trauma and addiction treatment for incarcerated men. *Journal of Anxiety Disorders* 2015; **30**(0): 66-80.
65. Rohde P, Jorgensen JS, Seeley JR, Mace DE. Pilot Evaluation of the Coping Course: A Cognitive-Behavioral Intervention to Enhance Coping Skills in Incarcerated Youth. *Journal of the American Academy of Child & Adolescent Psychiatry* 2004; **43**(6): 669-76.
66. Perkins R. The efficacy of mindfulness-based techniques in the reduction of stress in a sample of incarcerated women: Florida State University; 1998.
67. Richards JM, Beal WE, Seagal JD, Pennebaker JW. Effects of disclosure of traumatic events on illness behavior among psychiatric prison inmates. *Journal of Abnormal Psychology* 2000; **109**(1): 156-60.
68. Lanza PV, García PF, Lamelas FR, González-Menéndez A. Acceptance and Commitment Therapy Versus Cognitive Behavioral Therapy in the Treatment of Substance Use Disorder With Incarcerated Women. *Journal of Clinical Psychology* 2014; **70**(7): 644-57.
69. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis: Wiley; 2011.
70. Egger M, Smith GD, Altman D. Systematic reviews in health care: meta-analysis in context: John Wiley & Sons; 2008.
71. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in medicine* 2002; **21**(11): 1539-58.
72. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological methods* 2006; **11**(2): 193.
73. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal* 2003; **327**(7414): 557.
74. Wilson GL. Psychotherapy with Depressed Incarcerated Felons: A Comparative Evaluation of Treatments. *Psychological Reports* 1990; **67**(3): 1027-41.

75. Johnson JE, Zlotnick C. Pilot study of treatment for major depression among women prisoners with substance use disorder. *Journal of Psychiatric Research* 2012; **46**(9): 1174-83.
76. Ford JD, Chang R, Levine J, Zhang W. Randomized Clinical Trial Comparing Affect Regulation and Supportive Group Therapies for Victimization-Related PTSD With Incarcerated Women. *Behavior Therapy* 2013; **44**(2): 262-76.
77. Messina N, Grella CE, Cartier J, Torres S. A randomized experimental study of gender-responsive substance abuse treatment for women in prison. *Journal of Substance Abuse Treatment* 2010; **38**(2): 97-107.
78. Zlotnick C. Treatment of Incarcerated Women With Substance Abuse and Posttraumatic Stress Disorder, Final Report, 2002.
79. Zlotnick C, Johnson J, Najavits LM. Randomized Controlled Pilot Study of Cognitive-Behavioral Therapy in a Sample of Incarcerated Women With Substance Use Disorder and PTSD. *Behavior Therapy* 2009; **40**(4): 325-36.
80. Ahrens J, Rexford L. Cognitive Processing Therapy for Incarcerated Adolescents with PTSD. *Journal of Aggression, Maltreatment & Trauma* 2002; **6**(1): 201-16.
81. Abrams AI, Siegel LM. The Transcendental Meditation® Program and Rehabilitation At Folsom State Prison: A Cross-Validation Study. *Criminal Justice and Behavior* 1978; **5**(1): 3-20.
82. Valentine PV, Smith TE. Evaluating Traumatic Incident Reduction Therapy With Female Inmates: A Randomized Controlled Clinical Trial. *Research on Social Work Practice* 2001; **11**(1): 40-52.
83. Sleed M, Baradon T, Fonagy P. New Beginnings for mothers and babies in prison: A cluster randomized controlled trial. *Attachment & Human Development* 2013; **15**(4): 349-67.
84. Gussak D. The effects of art therapy on male and female inmates: Advancing the research base. *The Arts in Psychotherapy* 2009; **36**(1): 5-12.
85. Gold C, Assmus J, Hjørnevik K, et al. Music Therapy for Prisoners: Pilot Randomised Controlled Trial and Implications for Evaluating Psychosocial Interventions. *International Journal of Offender Therapy and Comparative Criminology* 2014; **58**(12): 1520-39.
86. Mitchell P, Smedley K, Kenning C, et al. Cognitive behaviour therapy for adolescent offenders with mental health problems in custody. *Journal of Adolescence* 2011; **34**(3): 433-43.
87. Bradley RG, Follingstad DR. Group therapy for incarcerated women who experienced interpersonal violence: A pilot study. *Journal of Traumatic Stress* 2003; **16**(4): 337-40.
88. Chandiramani K, Verma S, Dhar P. Psychological effects of Vipassana meditation on Tihar Jail inmates. Igatpuri, India: Vipassana Research Institute, 2000.
89. Chatterjee S, Hadi AS. Regression Analysis by Example: Wiley; 2006.
90. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Statistics in medicine* 2002; **21**(11): 1559-73.
91. Tacconelli E. Systematic reviews: CRD's guidance for undertaking reviews in health care. *The Lancet Infectious Diseases*; **10**(4): 226.
92. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials; 2011.
93. Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology* 2001; **54**(10): 1046-55.

94. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine* 2007; **26**(25): 4544-62.
95. Biggam FH, Power KG. A Controlled, Problem-Solving, Group-Based Intervention with Vulnerable Incarcerated Young Offenders. *International Journal of Offender Therapy and Comparative Criminology* 2002; **46**(6): 678-98.
96. Cole KL, Sarlund-Heinrich P, Brown LS. Developing and Assessing Effectiveness of a Time-Limited Therapy Group for Incarcerated Women Survivors of Childhood Sexual Abuse. *Journal of Trauma & Dissociation* 2007; **8**(2): 97-121.
97. Hilkey JH, Wilhelm CL, Horne AM. COMPARATIVE EFFECTIVENESS OF VIDEOTAPE PRETRAINING VERSUS NO PRETRAINING ON SELECTED PROCESS AND OUTCOME VARIABLES IN GROUP THERAPY. *Psychological Reports* 1982; **50**(3c): 1151-9.
98. Loper A, Tuerk E. Improving the Emotional Adjustment and Communication Patterns of Incarcerated Mothers: Effectiveness of a Prison Parenting Intervention. *J Child Fam Stud* 2011; **20**(1): 89-101.
99. Maunder L, Cameron L, Moss M, et al. Effectiveness of self-help materials for anxiety adapted for use in prison – a pilot study. *Journal of Mental Health* 2009; **18**(3): 262-71.
100. Pardini J, Scogin F, Schriver J, Domino M, Wilson D, LaRocca M. Efficacy and process of cognitive bibliotherapy for the treatment of depression in jail and prison inmates. *Psychological Services* 2014; **11**(2): 141-52.
101. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Archives of general psychiatry* 1961; **4**: 561-71.
102. Beck AT, Steer RA, Brown GK. Beck depression inventory-II. San Antonio, TX: Psychological Corporation; 1996.
103. Miller IW, Bishop S, Norman WH, Maddever H. The modified Hamilton rating scale for depression: Reliability and validity. *Psychiatry Research* 1985; **14**(2): 131-42.
104. Derogatis LR, Spencer P. Brief symptom inventory: BSI: Pearson Upper Saddle River, NJ; 1993.
105. Blake DD, Weathers FW, Nagy LM, et al. The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress* 1995; **8**(1): 75-90.
106. Spielberger CD. State-Trait Anxiety Inventory. The Corsini Encyclopedia of Psychology: John Wiley & Sons, Inc.; 2010.
107. Achenbach TM. Manual for the Child Behavior Checklist/4-18 and 1991 profile: Department of Psychiatry, University of Vermont Burlington, VT; 1991.
108. Derogatis LR, Unger R. Symptom Checklist-90-Revised. The Corsini Encyclopedia of Psychology: John Wiley & Sons, Inc.; 2010.
109. Schwartz G, Davidson R, Goleman D, Corcoran K, Fischer J. Cognitive somatic anxiety questionnaire (CSAQ). *K Corcoran, & J Fischer, Measures for clinical practice: A sourcebook* 2000: 128-9.
110. McLellan AT, Kushner H, Metzger D, et al. The fifth edition of the addiction severity index. *Journal of Substance Abuse Treatment* 1992; **9**(3): 199-213.
111. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *Journal of Psychosomatic Research* 2002; **52**(2): 69-77.
112. Herge W, Landoll R, La Greca A. Center for Epidemiologic Studies Depression Scale (CES-D). In: Gellman M, Turner JR, eds. Encyclopedia of Behavioral Medicine: Springer New York; 2013: 366-7.

113. Deacon BJ, Abramowitz JS, Woods CM, Tolin DF. The Anxiety Sensitivity Index - Revised: psychometric properties and factor structure in two nonclinical samples. *Behaviour Research and Therapy* 2003; **41**(12): 1427-49.
114. Bruss GS, Gruenberg AM, Goldstein RD, Barber JP. Hamilton anxiety rating scale interview guide: Joint interview and test-retest methods for interrater reliability. *Psychiatry Research* 1994; **53**(2): 191-202.
115. Westhuis D, Thyer BA. Development and Validation of the Clinical Anxiety Scale: A Rapid Assessment Instrument for Empirical Practice. *Educational and Psychological Measurement* 1989; **49**(1): 153-63.
116. Briere J. Trauma symptom inventory. *Psychological Assessment Resources, Odessa, Fla, USA* 1996.
117. Haden SC, Shiva A. A Comparison of Factor Structures of the Barratt Impulsiveness Scale in a Mentally Ill Forensic Inpatient Sample. *International Journal of Forensic Mental Health* 2009; **8**(3): 198-207.
118. Evans C, Connell J, Barkham M, et al. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE—OM; 2002.
119. Foa EB, Riggs DS, Dancu CV, Rothbaum BO. Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress* 1993; **6**(4): 459-73.
120. Cacciola JS, Alterman AI, Habing B, McLellan AT. Recent Status Scores for Version 6 of the Addiction Severity Index (ASI-6). *Addiction (Abingdon, England)* 2011; **106**(9): 1588-602.
121. Cohen J. Statistical power analysis for the behavioral sciences (rev. ed.). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc; 1977.
122. Huhn M, Tardy M, Spineli L, et al. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: A systematic overview of meta-analyses. *JAMA Psychiatry* 2014; **71**(6): 706-15.
123. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ : British Medical Journal* 2003; **327**(7414): 557-60.
124. Goff A, Rose E, Rose S, Purves D. Does PTSD occur in sentenced prison populations? A systematic literature review. *Criminal Behaviour and Mental Health* 2007; **17**(3): 152-62.
125. Wolff N, Huening J, Shi J, Frueh BC. Trauma Exposure and Posttraumatic Stress Disorder among Incarcerated Men. *J Urban Health* 2014; **91**(4): 707-19.
126. Green BL, Miranda J, Daroowalla A, Siddique J. Trauma Exposure, Mental Health Functioning, and Program Needs of Women in Jail. *Crime & Delinquency* 2005; **51**(1): 133-51.
127. Trestman RL, Ford J, Zhang W, Wiesbrock V. Current and Lifetime Psychiatric Illness Among Inmates Not Identified as Acutely Mentally Ill at Intake in Connecticut's Jails. *Journal of the American Academy of Psychiatry and the Law Online* 2007; **35**(4): 490-500.
128. Bond G, Evans L, Salyers M, Williams J, Kim H-W. Measurement of Fidelity in Psychiatric Rehabilitation. *Ment Health Serv Res* 2000; **2**(2): 75-87.
129. Williams MW, Skogstad P, Deane FP. Attitudes of Male Prisoners Toward Seeking Professional Psychological Help. *Journal of Offender Rehabilitation* 2001; **34**(2): 49-61.
130. Byrne P. Stigma of mental illness and ways of diminishing it; 2000.
131. Corrigan P. How Stigma Interferes With Mental Health Care. *American Psychologist* 2004; **59**(7): 614-25.
132. Meyer B, Pilkonis PA, Krupnick JL, Egan MK, Simmens SJ, Sotsky SM. Treatment expectancies, patient alliance and outcome: Further analyses from the National

Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology* 2002; **70**(4): 1051-5.
133. Ryan RM, Deci EL. A self-determination theory approach to psychotherapy: The motivational basis for effective change. *Canadian Psychology/Psychologie canadienne* 2008; **49**(3): 186-93.
134. Fazel S, Gulati G, Linsell L, Geddes JR, Grann M. Schizophrenia and Violence: Systematic Review and Meta-Analysis. *PLoS Med* 2009; **6**(8): e1000120.

Appendix A - PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	

Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	

Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

Appendix B – RCT Quality checklist used for quality appraisal

Section1: Internal Validity		
1.1	The study addresses an appropriate and clearly focused question.	Well covered Not addressed Adequately addressed Not reported Poorly addressed Not applicable
1.2	The assignment of subjects to treatment groups is randomised.	Well covered Not addressed Adequately addressed Not reported Poorly addressed Not applicable
1.3	An adequate concealment method is used.	Well covered Not addressed Adequately addressed Not reported Poorly addressed Not applicable
1.4	Subjects and investigators are kept 'blind' about treatment allocation.	Well covered Not addressed Adequately addressed Not reported Poorly addressed Not applicable
1.5	The treatment and control groups are similar at the start of the trial.	Well covered Not addressed Adequately addressed Not reported Poorly addressed Not applicable
1.6	The only difference between groups is the treatment under investigation.	Well covered Not addressed Adequately addressed Not reported Poorly addressed Not applicable
1.7	All relevant outcomes are measured in a	Well covered Not addressed

	standard, valid and reliable way.	Adequately addressed Not reported Poorly addressed Not applicable
1.8	What percentage of the individuals or clusters recruited into each treatment arm of the study dropped out before the study was completed?	Well covered Not addressed Adequately addressed Not reported Poorly addressed Not applicable
1.9	All the subjects are analysed in the groups to which they were randomly allocated (often referred to as intention-to-treat analysis).	Well covered Not addressed Adequately addressed Not reported Poorly addressed Not applicable
1.10	Where the study is carried out at more than one site, results are comparable for all sites.	Well covered Not addressed Adequately addressed Not reported Poorly addressed Not applicable
Section 2. Overall assessment of the study		
2.1	How well was the study done to minimise bias? Code ++ , + or –	Well covered Not addressed Adequately addressed Not reported Poorly addressed Not applicable

Appendix C – Study characteristics of included studies

Items not recorded in the study are left blank. Secondary outcome measures include only outcome measures that were examined in the analyses.
Abbreviations: N = sample size, wk. = week, hr = hour. For studies that did not provide mean age but report age range of the sample, the range was recorded.

Author & Year	Location	N	% Male	Age (Mean)	Diagnosis	Treatment type	Treatment length	Primary outcome	Secondary outcome measures
Chandiramani et al., 1995	India, Tihar Jail	150	100.00%	-	Not required	Vipassana Retreats (meditation-based)	10 days	BDI	Hamilton Anxiety Rating Scale
Valentine & Smith, 2001	USA, Florida Federal Prison	123	0.00%	32.8	Not required	Traumatic Incident Reduction Therapy	3-4 hours	BDI	Clinical Anxiety Scale (CAS), PTSD Symptom Scale (PSS)
Johnson & Zlotnick, 2012	USA, Rhode Island State Prison	38	0.00%	35	MDD & SUD	Interpersonal Psychotherapy	8 weeks	Hamilton Rating Scale for Depression (HRSD)	-
Ford et al., 2013	USA, York Correctional Institution (Connecticut state prison)	72	0.00%	34.6	Full/partial PTSD	Trauma Affect Regulation: Guide for Education and Therapy (TARGET)	Twelve 75-minute sessions	Clinician Administered PTSD Scale (CAPS) - Total PTSD Symptoms	Trauma Symptom Inventory (TSI), Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM)

Gold et al., 2014	Norway, Bergen Prison	113	100.00%	31.38	Not required	Music Therapy	4 sessions median (5.27 sessions average)	STAI-State (State Anxiety)	STAI-Trait, HADS-A (Anxiety), HADS-D (Depression)
Zlotnick, 2002a	USA, Rhode Island Adult Correctional Institution	28	0.00%	30.9	SUD & PTSD	Seeking Safety	12 weeks: 90 min-twice/wk.	Clinician Administered PTSD Scale (CAPS-I)	Addiction Severity Index (ASI)
Mitchell et al., 2011	USA, Secure Children's Homes, Young Offender Institution	40	100.00%	15.58	Not required but included diagnosed sample	CBT	9.1 sessions mean - weekly session	Youth Self Report (YSR) - Total behavioural & emotional functioning	-
Khodayarifard et al., 2010	Iran, Rajaee Shahr Prison	86	100.00%	48.23	Not required	CBT (Combined)	Avg. (8) 1-hr weekly group + (8) 1-hr weekly individual sessions	SCL-90-R (Global Severity Index)	General Health Questionnaire (GHQ-28)
Khodayarifard et al. (2), 2010	Iran, Rajaee Shahr Prison (Tehran province)	88	100.00%	48.23	Not required	CBT (Individual)	16 weeks: 2-hr weekly	SCL-90-R (Global Severity Index)	General Health Questionnaire (GHQ-28)
Wilson, 1990	USA, large maximum-security prison	10	100.00%	33.1	Depression	Group cognitive therapy	Fourteen 90-min sessions (21 hrs)	BDI	-

Bradley & Follingstad, 2003	USA, south-eastern state medium-security prison	31	0.00%	36.67	Trauma symptom and/or Depression	Trauma treatment group therapy (DBT-based)	18 sessions, 9- 2.5 hrs, 9-writing	BDI	Trauma Symptom Inventory (TSI)
Gussak, 2007a	USA, Florida, medium- to maximum security prisons	29	100.00%	21–59	Not required; but many had Axis I Disorders	Art Therapy	8 weeks: weekly group session	BDI-II	-
Gussak, 2009b	USA, Florida maximum security prison	60	100.00%	22–51	Not required	Art Therapy	15 weeks, weekly	BDI-II	-
Gussak (2), 2009b	USA, Florida medium-security prison	96	0.00%	20–51	Not required	Art Therapy	15 weeks, weekly	BDI-II	-
Abrams & Siegal, 1979	USA, CA Folsom State Prison	55	100.00%	-	Not required	Transcendental Meditation	14 weeks, weekly	STAI – State Anxiety	STAI- Trait Anxiety
Abrams & Siegal (2), 1979	USA, CA Folsom State Prison	49	100.00%	-	Not required	Transcendental Meditation	14 weeks, weekly	STAI – State Anxiety	STAI-Trait
Hilkey & Horne, 1982	USA, Indiana medium-security Federal Penitentiary	89	100.00%	28.1	Not required	Video pre-training on Group psychotherapy	8 weeks: weekly 90min (30-min pre-training)	STAI – State Anxiety	-

Maunder et al., 2009	UK, North England, HMP Acklington	38	100.00%	35.22	Symptoms of Anxiety, Depression	Self-help materials	4 weeks	Hospital Anxiety and Depression Scale (HADS)	BSI
Richards et al., 2000	USA, Midwestern max-security psychiatric prison	98	100.00%	34.5	>1 DSM-III-R diagnosis	Writing about trauma	3 days-20 min/day	Cognitive Somatic Anxiety Questionnaire (CSAQ)	-
Zlotnick et al., 2009b	USA, minimum security wing of women's prison	49	0.00%	34.6	PTSD & SUD	Seeking Safety	6-8 wks.: 90min, 3 sessions per/week; after release 60-min weekly booster for 12 weeks	CAPS-I	BSI
Pardini et al., 2013	USA, Tuscaloosa, Alabama metropolitan jail	37	78.40%	29.37	Depression (BDI score ≥ 10)	Bibliotherapy	4 weeks	BDI	SCL-90-R (Global Severity Index)
Pardini et al. (2), 2013	USA, Alabama, max security prison	42	100.00%	32.7	Depression (CES-D score ≥ 10)	Bibliotherapy	4 weeks	BDI-II	HRSD, BSI total
Perkins, 1998	USA, Florida, Tallahassee, Federal Correctional Institution	143	0.00%	33.9	Not required	Meditation based program	8 weeks: 2hr/wk.	SCL-90-R (Global Severity Index)	Coping Resources Inventory for Stress (CRIS) – physical health subscore

Messina et al., 2010	USA,CA Valley State Prison for Women	115	0.00%	36	Not required	Gender-responsive treatment (GRT)	17 sessions of *Helping Women Recover* + 11 sessions of *Beyond Trauma*	ASI – psychological composite score	ASI – other subscores
Biggam & Power, 2002	Scotland, Young Offenders Institution	46	100.00%	19.3	Not required	Problem solving group-based intervention	Five 90-min sessions	HADS Depression	HADS Anxiety
Ahrens & Rexford, 2002	USA, youth correctional facility	38	100.00%	16.4	PTSD	CPT (Cognitive Processing Therapy)	8 weeks: eight 60-min sessions	BDI	PTSD Symptom Scale-Self Report (PSS-SR)
Sumter et al., 2009	USA, Tidewater Detention Center, Chesapeake, Virginia	33	0.00%	-	Not required	Meditation	7 weeks, 2.5 hr/wk.	Sleeping difficulties – symptom	Emotions – Hopelessness scale
Bilderbeck et al., 2013	UK, 7 prisons	167	92.81%	36.68	Not required	Yoga	10 weeks	BSI	Perceived Stress Scale (PSS)
Chen, 2014	China, Beijing, adult male prison	194	100.00%	35.5	Anxiety level $\geq$ 49 on STAI or depression level $\geq$ 14 on BDI	Group Music Therapy	20 session: 90-mi, twice weekly	BDI	STAI

Sleed et al., 2013	UK, MBU (Mother Baby Units) in Prisons	153	0.00%	-	Not required	New Beginnings (Attached-based group intervention)	4 weeks: (8 sessions) 2 session of 2hr/day, 1day/wk.	Center for Epidemiologic Studies Depression Scale (CES-D)	-
Villagra Lanza et al. (1), 2014	Spain, Module 10, Prison of Villabona (Asturias)	31	0.00%	32	SUD	ACT (Acceptance and Commitment Therapy)	16 weeks: weekly 90-min sessions	Anxiety Sensitivity Index (ASI)-Total	-
Villagra Lanza et al. (2), 2014	Spain, Villabona Prison	50	0.00%	33.2	SUD	CBT	16 weeks: weekly 90-min sessions	Anxiety Sensitivity Index (ASI) – Total	ASI-6 subscores
Cole et al., 2007	USA, Washington Correction Center for Women	9	0.00%	31	Childhood Sexual Abuse	Time-Limited Trauma-focused Therapy Group	8 weeks: 16 sessions, biweekly, 2.5 hrs	SCL-90-R (Global Severity Index)	Trauma symptom inventory
Loper & Tuerk, 2011	USA, State prison	106	0.00%	33	Not required	Parenting Intervention	8 sessions	BSI – Global Severity Index	BSI subscores
Wolff et al., 2015	USA, Pennsylvania, high security prison	173	100.00%	42.5	Not required	Seeking Safety	14 weeks	CAPS score (mean)	BSI – Global Severity Index (mean)

Wolff et al.(2), 2015	USA, Pennsylvania, high security prison	175	100.00%	-	Not required	Male-Trauma Recovery Empowerment Model (M-TREM)	14 weeks	CAPS score (mean)	BSI – Global Severity Index (mean)
Rohde et al., 2004	USA, Oak Creek Youth Correctional Facility	76	100%	16.30	Not required	Coping Course	8 weeks: 2 times/wk.	YSR – internalizing score	-

Appendix D – Quality rating of included studies (using quality checklist from Appendix B)

Full description of quality checklist questions can be found in Appendix B. Abbreviations: equiv. = equivalent, sig. = significant, expr = experimental.
Key for rating: 1 = Well covered, 2 = Adequately addressed, 3 = Poorly addressed, N/A = Not addressed/ not applicable.
++ = **all or most** of the criteria fulfilled (unfulfilled criteria very unlikely to alter conclusions of study)
+ = **some** of the criteria fulfilled (unfulfilled criteria unlikely to alter conclusions of study).

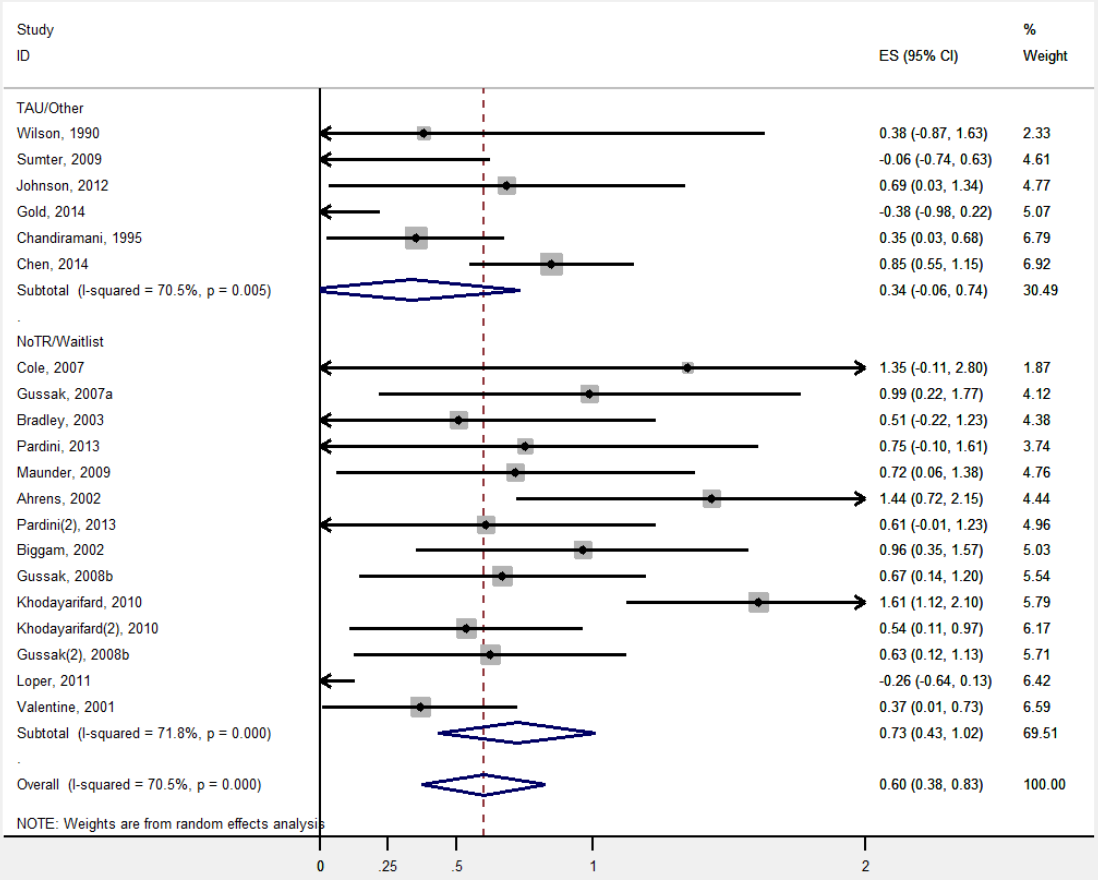
Author & Year	1.1 question	1.2 random	3 concealment	4 double-blind	5 similar group	6 controlled	7 standard outcome measure	8 attrition rate	9 ITT (Intention-to-Treat analysis)	10 multiple site equiv.	2.1 overall
Chandiramani et al., 1995	1	2	N/A	N/A	2 - more illiterate in experimental group	2	1	0% (not explicitly stated)	1	1	+
Valentine & Smith, 2001	1	1	2	N/A	1	1	1	17%	1	1	++
Johnson & Zlotnick, 2012	1	1	1	1	1	1	1	11%	1	1	++
Ford et al., 2013	1	1	1	1	1	1	1	10%	2	1	++
Gold et al., 2014	1	1	N/A	N/A	1	1	1	39%	1	1	+
Zlotnick, 2002a	1	1	N/A	N/A	1	1	1	0%	N/A	1	+
Mitchell et al., 2011	1	1	1	1	1	1	1	5%	1	1	++
Khodayarifard et al., 2010	1	1	N/A	N/A	1	1	1	28%	N/A	1	+

Khodayarifard et al.(2), 2010	1	1	N/A	N/A	1	1	1	27%	N/A	1	+
Wilson, 1990	1	1	2	N/A – same therapist for all conditions	1	1	1	0%	1	1	+
Bradley & Follingstad, 2003	1	1	N/A	N/A	1	1	1	37%	1	1	+
Gussak, 2007a	1	1	2	2	3 – at beginning of experiment: BDI score higher in expr group)	3	1	41%	1	1	+
Gussak, 2009b	1	1	N/A	N/A	1	1	1	48%	N/A	1	+
Gussak (2), 2009b	1	1	N/A	N/A	1	1	1	24%	N/A	1	+
Abrams & Siegal, 1979	1	1	N/A	N/A	1	1	1	27%	1	1	+
Abrams & Siegal (2), 1979	1	1	N/A	N/A	1	1	1	17%	1	1	+
Hilkey & Horne, 1982	1	1	N/A	2	2	1	1	0%	1	1	+
Maunder et al., 2009	1	1	1	1 – blind researcher	1	1	1	50%	1	1 – pilot study	+
Richards et al., 2000	1	1	N/A	N/A	1	1	1	7%	N/A	1	+
Zlotnick et al., 2009b	1	1	N/A	N/A	1	1	1	10%	1	1	+

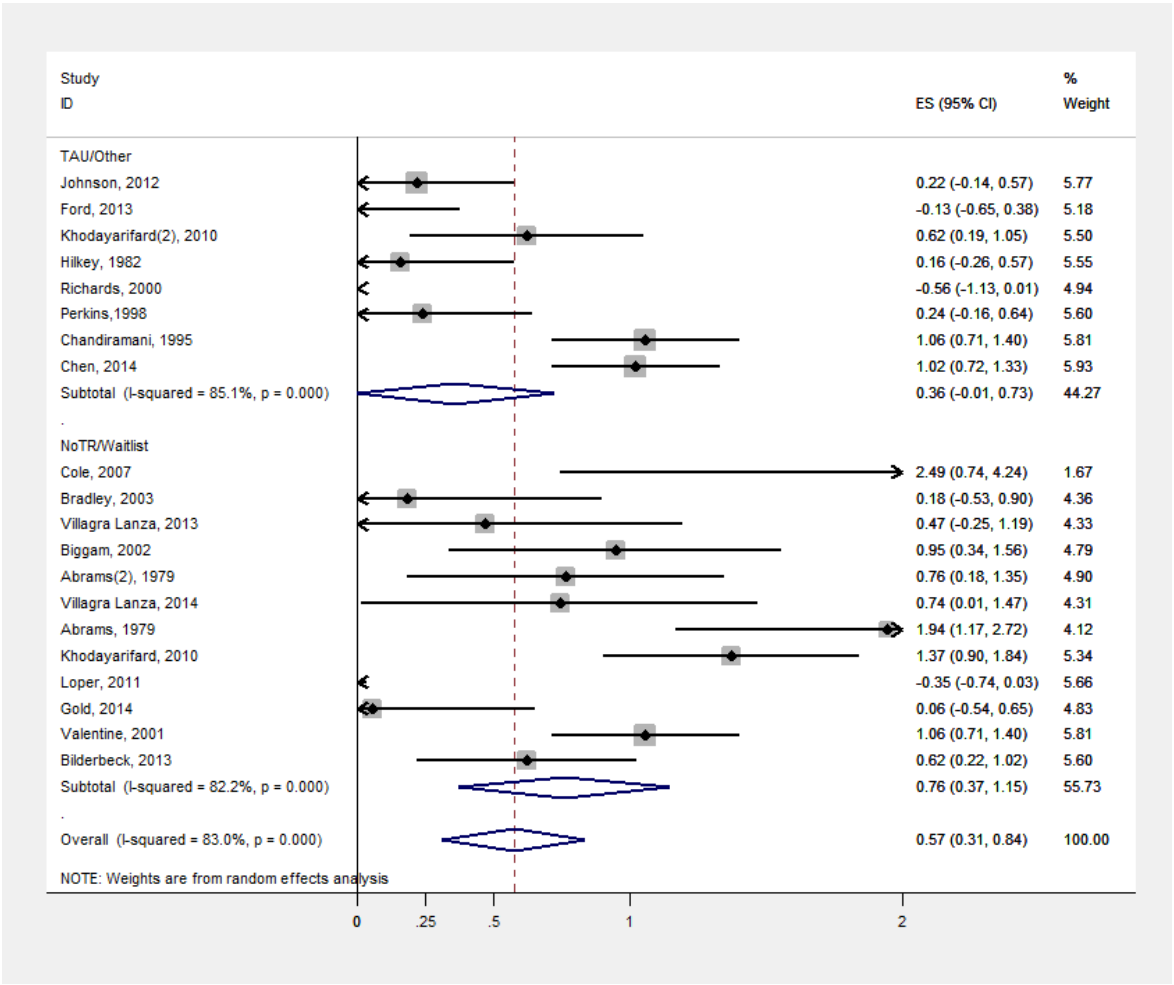
Pardini et al., 2013	1	1	1	N/A	1	1	1	38%	1	1 – pilot study	+
Pardini et al. (2), 2013	1	1	1	1 – blind researcher	1	1	1	17%	1	1	++
Perkins, 1998	1	1	N/A	N/A	1	1	1	10%: 15 drop out; rest of the attrition due to transfers, WRIT, release	N/A	1	+
Messina et al., 2010	1	1	1	N/A	1	1	1	18%	1	1	++
Biggam & Power, 2002	1	1	2	2	1	2	1	41%	1	1	+
Ahrens & Rexford, 2002	1	1	N/A	N/A	2	1	1	0% (not explicitly stated)	1	1	+
Sumter et al., 2009	1	1	N/A	N/A	1	1	3	0% (not explicitly stated)	1	1	+
Bilderbeck et al., 2013	1	1	N/A	N/A	1	1	1	31%	N/A	1	+
Chen, 2014	1	1	1	N/A	1	1	1	8%	1	1	+
Sleed et al., 2013	1	1	1	1	2	1	1	29%	1	1	++
Villagra Lanza et al. (1), 2014	1	1	N/A	N	2	2	1	19% at follow-up; 0% at post-test	1	1	+

Villagra Lanza et al. (2), 2014	1	1	N/A	N/A	1	1	1	3%	1	1	+
Cole et al., 2007	1	1	N/A	N/A	1	2	1	31%	1	1	+
Loper & Tuerk, 2011	1	1	N/A	N/A	1	2	1	18%	1	1	+
Wolff et al., 2015	1	2 – random & preference groups; no sig. difference between two groups	N/A	N/A	1	2	1	13%	1	1	+
Wolff et al.(2), 2015	1	2 – random & preference groups; no sig. difference between two groups	N/A	N/A	1	2	1	10%	1	1	+
Rohde et al., 2004	1	1	N/A	N/A	2	2	1	0%	1	1	+

Appendix E – Effect sizes of studies with depression outcomes (by comparator type).
TAU/other include other types of psychological treatment. NoTR = no treatment.

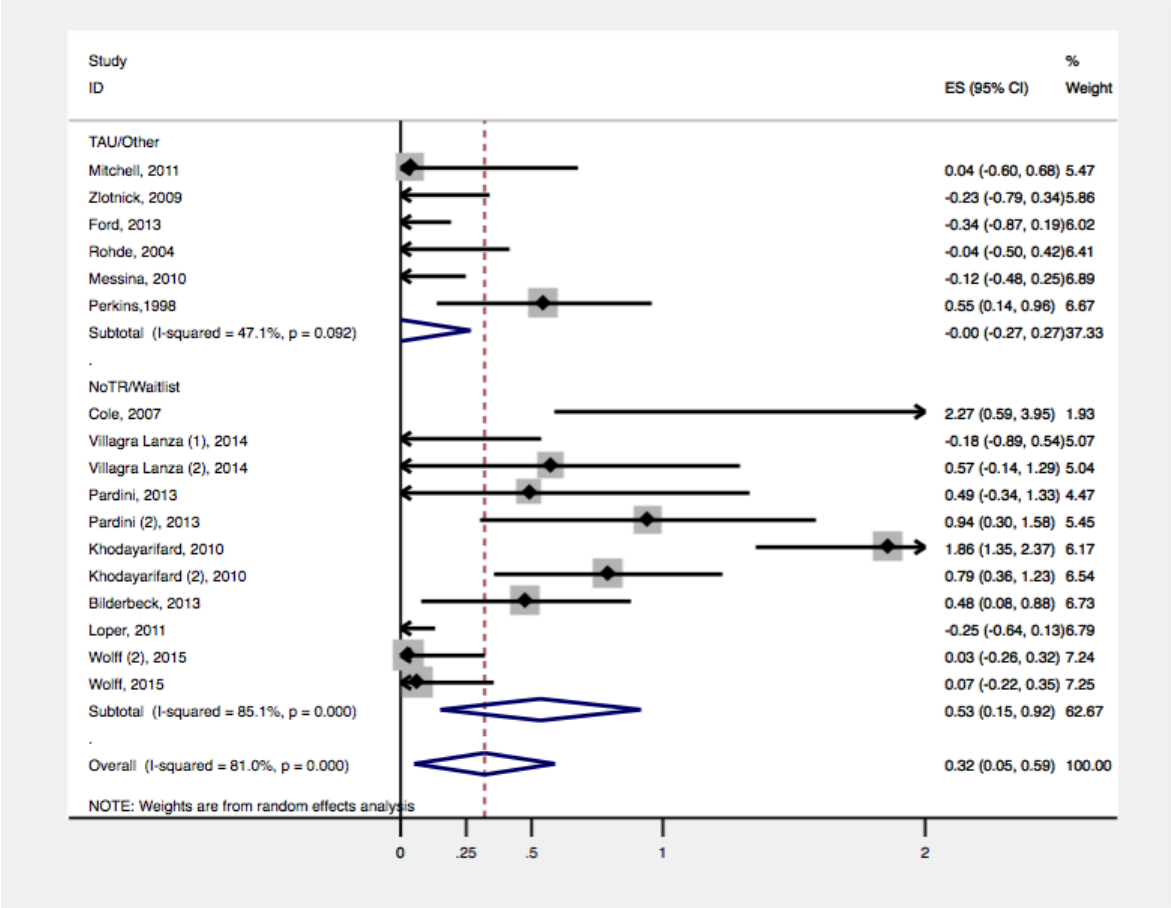


Appendix F – Effect sizes of studies with anxiety outcomes (by comparator type).
TAU/other include other types of psychological treatment. NoTR = no treatment.

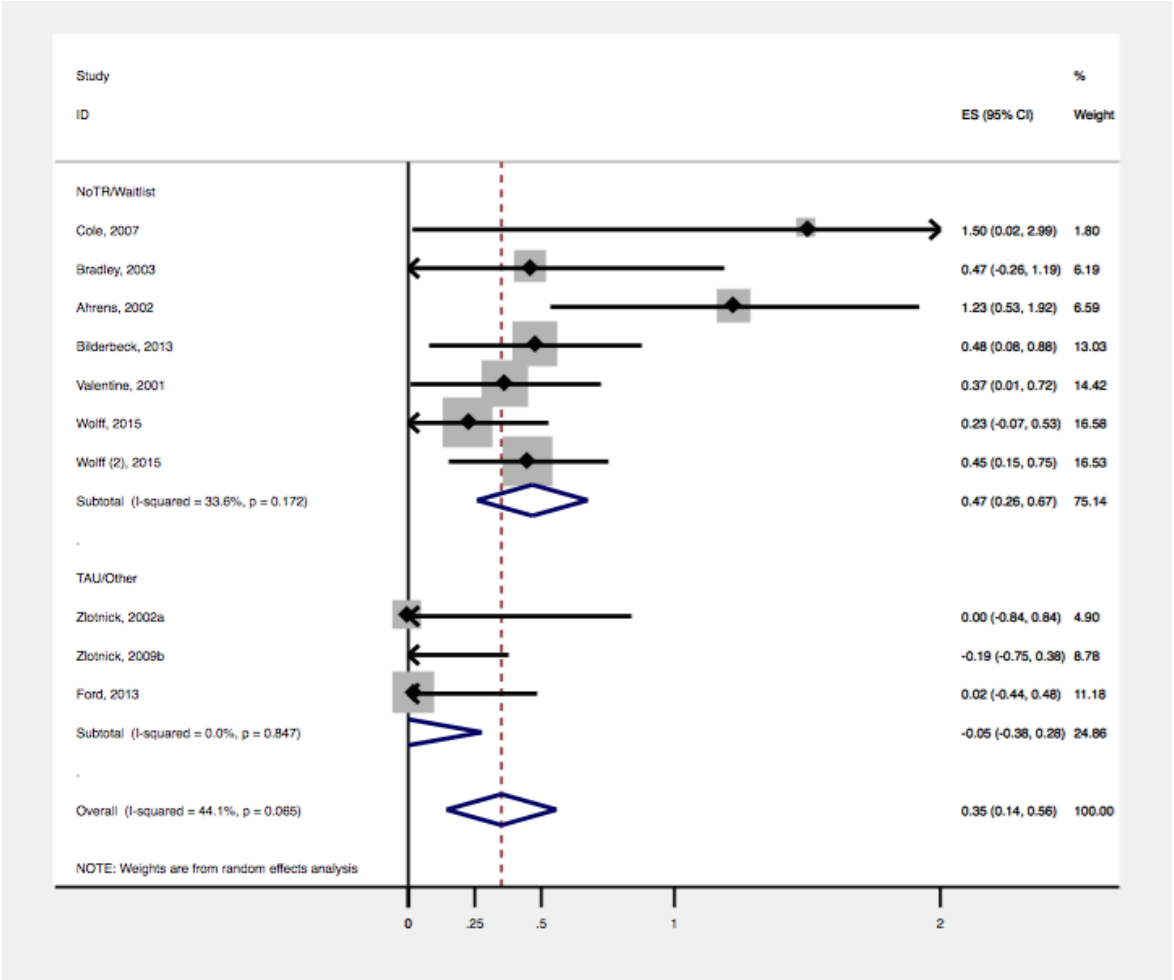


Appendix G – Effect sizes of studies with overall psychopathology outcomes (by comparator type)

TAU/other include other types of psychological treatment. NoTR = no treatment.

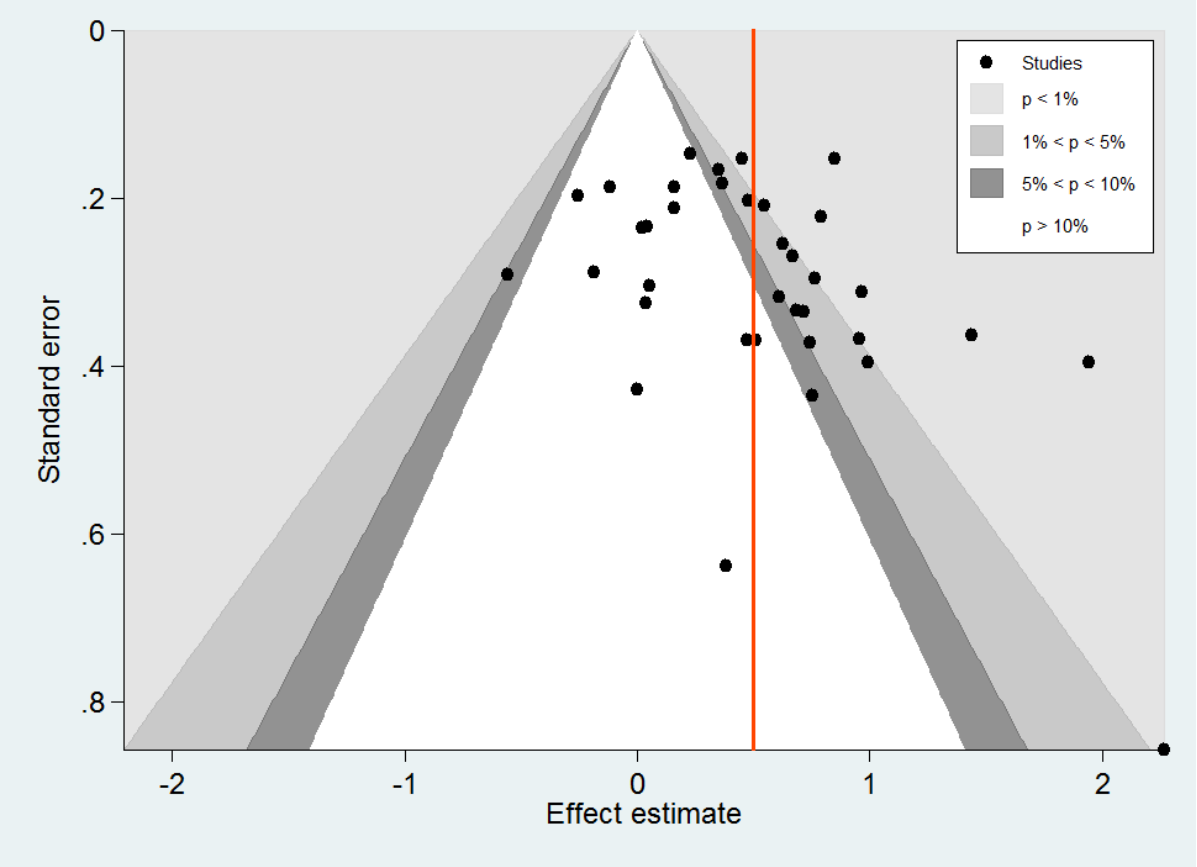


Appendix H – Effect sizes of studies with trauma outcomes (by comparator type).
TAU/other include other types of psychological treatment. NoTR = no treatment.



Appendix I – Contoured funnel plot of total studies.

Red axis in the centre indicates observed pooled effect estimate.



Appendix J – Contoured funnel plot of studies with no treatment/waitlist controls

Red axis in the centre indicates observed pooled effect estimate.

