Effectiveness of psychological and psychosocial interventions for forensic mental health inpatients

Effectiveness of psychological and psychosocial interventions for forensic mental health inpatients: A meta-analysis

Article *in* Aggression and Violent Behavior · December 2020

DOI: 10.1016/j.avb.2021.101551

CITATIONS
0

READS
31

4 authors:

Lindsey Gilling McIntosh
The University of Edinburgh
14 PUBLICATIONS **247** CITATIONS
SEE PROFILE

Sarah Janes
The University of Edinburgh
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Suzanne O'Rourke
The University of Edinburgh
31 PUBLICATIONS **874** CITATIONS
SEE PROFILE

Lindsay Thomson
The University of Edinburgh
110 PUBLICATIONS **652** CITATIONS
SEE PROFILE

Effectiveness of psychological and psychosocial interventions for forensic mental health inpatients: A meta-analysis

Lindsey Gilling McIntosh [a,b], Sarah Janes [c], Suzanne O'Rourke [b,c], Lindsay D. G. Thomson [a,b]

[a] University of Edinburgh, Division of Psychiatry

[b] NHS State Hospitals Board for Scotland

[c] University of Edinburgh, Department of Clinical Psychology

Corresponding author:

Lindsey Gilling McIntosh, l.m.gilling-mcintosh@sms.ed.ac.uk

University of Edinburgh Division of Psychiatry, Royal Edinburgh Hospital, Morningside Place, Edinburgh

Telephone: (+44) 131 537 6000

Declarations of interest: None declared.

Funding: This work was supported by The State Hospitals Board for Scotland Ref: G33448 Effectiveness of the Forensic Matrix. The funder did not have a role in the study design, collection, analysis or interpretation of the data, writing of the manuscript or in the decision where to submit the paper for publication.

Abstract

Psychological and psychosocial interventions have significant potential to treat the mental health and criminogenic needs of forensic mental health patients. However, due to a dearth of high-quality evaluation studies there is limited evidence on the effectiveness of these interventions with this population. This review aimed to quantitatively summarise the effectiveness of psychological therapies delivered in forensic hospitals. The literature was systematically searched to identify controlled evaluations of psychological interventions delivered within inpatient forensic psychiatric settings. Twenty-eight studies were included. Methodological quality was assessed using the SIGN Methodology Checklists. Pooled effect sizes were calculated for fourteen outcome domains. Small effect sizes were found favouring psychological treatment over the comparator condition in increasing insight into mental illness, ameliorating symptoms, improving problem-solving ability, reducing pro-criminal attitudes and improving ward behaviour. A medium effect size was found for treatment increasing patients' knowledge of their mental illness. There were few outcomes for which psychological therapy was associated with improvements beyond that of comparison treatment, and these improvements were generally small. Despite more frequent adoption of a randomised-controlled design, methodological quality remains problematic and more well-designed trials are needed to determine the effectiveness of psychological interventions across outcome domains relevant to forensic patients' recovery.

Keywords

Review, meta-analysis, effectiveness, forensic, psychological therapies

1. Introduction

Individuals detained in the care of forensic mental health services often present with entrenched, complex mental health problems alongside a range of offending behaviours. By working with forensic patients, forensic mental health practitioners aim to assess and treat patients' mental health needs and reduce their risk of future reoffending while promoting patients' personal recovery (Barnao & Ward, 2015; Vojt, Slesser, Marshall & Thomson, 2011). Treatment in forensic hospitals is often long-term and comes with great financial and personal cost. In England, where secure mental health services account for nearly 20% of all public expenditure on adult mental healthcare (Durcan, Hoare, & Cumming, 2011), the average length of stay in continuous secure care in a sample of high and medium secure patients was over 14 years (Völlm et al., 2018). Considering this significant investment of resources, there is a dearth of empirical evidence for the interventions delivered in forensic mental health settings, particularly psychological interventions (Barnao & Ward, 2015; Mallion, Tyler & Miles, 2019). With few psychological or psychosocial interventions supported by research evidence in this population, the day-to-day treatment delivered is rarely evidence-based. A recent national survey of UK forensic hospitals found that a majority of the offence-specific group treatments delivered by services including substance misuse, firesetting, sexual offending or violent offending treatment programmes, were developed "in house' and only a minority had ever been subject to even a local service-evaluation of treatment effectiveness (Mallion, Tyler, & Miles, 2019).

Determining the effectiveness of interventions for forensic patients and services to deliver interventions found to be effective is crucial if healthcare providers are to remain committed to delivering evidence-based care. There have been many attempts to summarise and synthesise the research evidence for psychological treatment of forensic mental health patients and mentally disordered offenders (e.g. Duncan et al., 2006; MacInnes & Masino, 2019; Sturgeon, Tyler, & Gannon, 2018). Most commonly this includes systematic reviews followed by narrative synthesis, rather than quantitative synthesis via meta-analysis. Authors of narrative reviews of the literature on psychological treatment for forensic patients reported their initial plans to undertake meta-analysis were thwarted by a high prevalence of small scale

studies with lax methodological design (Duncan et al., 2006) and high levels of heterogeneity reflected in the treatments under evaluation and the outcomes being assessed (MacInnes & Masino, 2019).

A small number of meta-analyses with relevance to treatment in secure settings have been conducted (e.g. Martin et al., 2012; Papalia et al., 2019; Yoon, Slade, & Fazel, 2017) but their broad inclusion of studies on prisoners, forensic community outpatients as well as inpatients reduces the generalisability of review findings to patients in long-term forensic inpatient care. Firstly, factors such as the therapeutic skill-set of the intervention facilitator (e.g. clinician vs. correctional officer) and practical issues arising from the treatment environment may impact the effectiveness of treatments delivered in hospital compared to prison or community settings (for example, see Taylor et al., 2020). Secondly, aside from the obvious differences in violence risk between those detained in institutions and those residing in the community, many distinctions can be made between mentally disordered offenders (MDOs) in prisons and secure hospitals (Thomas, McCrone, & Fahy, 2009) which could conceivably lead to differential effectiveness of the same treatment. For example, Thomas et al. (2009) compared prisoners attending the prison mental health clinic to a sample of forensic inpatients. Forensic inpatients were more likely to have a diagnosis of a psychotic disorder, a history of drug misuse, and a history of previous psychiatric admissions than prisoners, while prisoners tend to have more unmet psychiatric and daily living needs. Other research has linked forensic patients' problems with impulsivity and behavioural control (Cullen et al., 2011) and ongoing psychotic symptoms and cognitive impairment to treatment dropout from interventions originally developed for use in correctional settings. A widely studied cognitive skills programme is Reasoning and Rehabilitation (R&R). There is a substantial evidence base supporting the effectiveness of R&R in reducing recidivism among violent offenders in community and institutional settings across the world (Tong & Farrington, 2006). However, there are striking examples of implementation failures of R&R in forensic psychiatric settings. For example, Cullen et al. (2012a; 2012b) found only 50% of forensic patients who started the original 36-session R&R completed treatment. R&R was subject to a significant revision to increase its responsivity to forensic patients' needs, including a reduction in the number of sessions from 36 to 16, the addition of a module on cognitive impairments, and

the introduction of individual mentoring between group sessions (Young, Chick & Gudjonsson, 2010). Yet, trialling the new "R&R2M", Young et al. (2010) reported 19% of patients who were referred to the group refused to attend, another 22% of patients who initially agreed subsequently failed to start the programme and a further 21% of patients started the R&R2M but terminated treatment prematurely. For the above reasons reviews which adopt broad inclusion criteria but do not undertake moderator analyses exploring the potential for differential treatment effectiveness dependent on the institutional setting or other study characteristics may be of little relevance for practitioners seeking to identify and introduce effective psychological interventions in forensic hospitals.

According to systematic reviews the most effective psychological treatments in secure hospitals appear to be those with an evidence-base in general psychiatry (Dumont et al., 2018). Narrative syntheses (MacInnes & Masino, 2019; Sturgeon, Tyler, & Gannon, 2018; Tapp et al., 2013) have found that psychoeducational programmes lead to improved insight and understanding of one's mental disorder. Cognitive behavioural therapy (CBT), largely considered the gold standard treatment for schizophrenia, depression and anxiety, as well as personality disorder (Hofmann et al., 2012) also reduces forensic patients' psychiatric symptoms. A range of therapies have shown promise in reducing violence risk and aggressive behaviour, including cognitive skills programmes (notwithstanding the implementation problems already noted), dialectical behavioural therapy (DBT) and CBT (MacInnes & Masino, 2019; Sturgeon et al., 2018). Despite the range of reviews undertaken to summarise this literature and a surge in the number of publications of controlled evaluations of psychological interventions which focus on this patient group, to date there has been no meta-analysis of the effectiveness of psychological interventions in addressing forensic mental health inpatients' needs arising from their mental illness and offending behaviours. A meta-analysis would provide a necessary foundation for objective consideration of psychological treatment effectiveness and would facilitate comparisons based on effectiveness and cost-effectiveness with other treatment modalities targeting similar outcomes, including psychotropic medication. This review therefore addressed this gap by conducting a meta-analytic review of controlled evaluations of psychological and psychosocial treatments delivered in forensic mental health hospitals.

2 Method

2.1 Protocol registration

This review conforms to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009). The protocol for this review was published with the PROSPERO register on 5 March 2016 (registration number: CRD42016036052).

2.2 Search strategy

The search strategy was designed to be highly sensitive and yield a high degree of recall. PsycINFO, MEDLINE, EMBASE and CINAHL Plus databases were searched from 1980 until 7 July 2020. Cochrane Central Register of Controlled Trials, and the International Standard Randomised Controlled Trials Registry (ISRCTN) were searched to identify registered protocols of controlled and randomised controlled trials, respectively. Finally, ProQuest Theses and Dissertations and OpenGrey databases were searched to identify unpublished evaluations.

The search strategy included MeSH terms, exploded terms, and text words to avoid missing articles that were incorrectly coded. This resulted in a high degree of recall, increasing the likelihood of identifying all relevant papers. A similar search strategy was applied to each database, though the precise search terms varied for each database depending on available subject headings and search fields. A complete list of search terms used for PsycINFO database is included in Appendix A. Searches were restricted to articles published in 1980 and afterward, as well as those indexed as researching adults (ages 18+). The year 1980 was selected as the initial search year because it aligns with the release of the Diagnostic and Statistical Manual of Mental Disorders-III (DSM-III) and the resulting broad changes to the criteria for many mental disorders.

Reference lists from previous relevant reviews (Duncan et al., 2006; MacInnes & Masino, 2019; Ross, Quayle, Newman, & Tansey, 2013; Sturgeon, Tyler, Gannon, 2018; Tapp, Perkins, Warren, Fife-Schaw, & Moore, 2013) and meta-analyses (Martin et al., 2012; Morgan et al., 2012) on treatments for

mentally disordered offenders were reviewed for potential eligible studies. The table of contents for *Criminal Behaviour and Mental Health* (from 1996 volume 6 to 2020 volume 30(2)), *International Journal of Forensic Mental Health* (from 2002 volume 1 to 2020 volume 19(2)), *The Journal of Forensic Psychiatry* (from 1990 Volume 1 to 2002 volume 13), and *The Journal of Forensic Psychiatry and Psychology* (from 2003 volume 14 to 2020 volume 31(3)) were also reviewed for potential studies for inclusion. Finally, the reference lists of all included studies were also reviewed to identify studies that may have been missed by other search methods.

2.3 Eligibility criteria

Inclusion and exclusion criteria were defined as follows, using the PICO framework:

Population: Studies were included if the sample was comprised of adults (age 18+), diagnosed with serious mental disorders, and inpatients in a secure therapeutic environment. This included both forensic mental health units as well as designated mental health wards in prisons or jails. Studies were excluded from the review if the study sample primarily comprised individuals with a primary diagnosis of learning disability or with an IQ below 70, or individuals diagnosed with substance use disorders without a co-occurring major mental illness.

Intervention: Studies were included if they reported an evaluation of a psychological or psychosocial intervention following a pre-defined structure (e.g. delivered using a manual; sessions organized by topics or modules) and intended to address patients' mental health needs or needs relating to offending behaviour. Therapy could be delivered in group or on individual basis, or a combination. Evaluations of arts therapy (dance, music, art) or animal-assisted therapies were excluded.

Comparison: Studies were included if reporting the results of a randomised controlled, non-randomised controlled (quasi-experimental) design or a controlled before and after design (which included non-contemporaneous control conditions). Any comparator treatment condition was eligible, including active comparison treatments or treatment as usual (TAU).

Outcomes: Studies were included if treatment outcomes included patient-related quantitative measures collected at pre & post intervention relating to psychosocial (e.g. psychological distress) and/ or offending-behaviour outcomes (e.g. recidivism). Studies were excluded if no quantitative outcomes were reported.

2.4 Study selection and data extraction

All abstracts and titles were screened by the primary reviewer (author LGM). Studies that were clearly irrelevant to the review were excluded at this stage. Full-text articles were reviewed against the inclusion criteria by the primary reviewer and a random 25% sample of full-text articles were independently reviewed by a second reviewer (author SJ). Studies that met inclusion criteria were classified by study design using the Scottish Intercollegiate Guidelines Network (SIGN) study design algorithm (SIGN, 2020a). Full-text was unavailable for five results, four of which were unpublished postgraduate dissertations. Full-text articles not written in English were translated by native speakers of the relevant language (Dutch, German, and Japanese). A bespoke data extraction form was used to record information from all included studies. Extracted information included details of the study design and sample, intervention(s), measures used and the quantitative outcomes reported.

2.5 Quality Assessment

Study quality was assessed using the SIGN Methodology checklist for controlled trials (SIGN, 2020b). The SIGN controlled trials checklist considers studies on randomisation, allocation concealment, blinding, group equivalency, and analysis and reporting of outcomes. One point was assigned for each checklist criterion that was met; no point was assigned if the criterion was not met or there was insufficient information available. A criterion was scored 'not applicable' and assigned 1 point if it did not apply to the study and was not counted against the study. Each study was rated out of a maximum of 10 points: ≥ 9 indicated high quality, 7-8 indicated acceptable quality, 6 or below indicated low quality. Quality

assessment was completed independently by two reviewers for all eligible studies. Disagreements in ratings were reviewed and discussed by the two reviewers until a consensus was agreed.

2.6 Meta-analysis

Studies were grouped in preparation for meta-analysis according to the treatment outcomes reported. Studies need not have used the same measurement tool to be grouped by similar outcome. Meta-analysis was carried out for each outcome for which there was at least three primary studies (*k*) reporting treatment effects. Meta-analyses were therefore conducted on the following treatment outcomes: insight into mental illness and attitudes toward treatment ($k = 5$), knowledge about mental illness ($k = 4$), psychiatric symptoms ($k = 11$), quality of life ($k = 3$), self-esteem ($k = 3$), coping skills ($k = 4$), internal locus of control ($k = 3$), empathy ($k = 4$), problem-solving ability ($k = 9$), impulsivity ($k = 3$), ward behaviour ($k = 8$), criminal and violent attitudes ($k = 9$), self-reported anger or aggression ($k = 10$) and institutional violence and aggression ($k = 3$). Between-group effect sizes (standardised mean difference) were computed from post-intervention group means and standard deviations. For studies that included an additional post-treatment follow up assessment, only outcomes from the first post-treatment assessment were used. Rarely did studies report having undertaken follow-up assessments some length of time after the treatment ended, and it was not possible to carry out meta-analysis with this small number of studies.

Standard deviations were reconstructed from reported *p*-values or *t*-statistics when necessary. Intention-to-treat outcomes were used when reported by the study. For studies that reported several outcomes per outcome type (e.g. two psychiatric symptoms measures), these were pooled to result in an average effect size for the outcome domain prior to meta-analysis. In two studies (Aho-Mustonen et al., 2011; Doyle et al., 2016) post-intervention data were not reported so the between group effect size was calculated from the reported pre-post treatment change scores.

Meta-analysis was carried out in R (R Core Team, 2020) using the *metafor* package (Viechtbauer, 2010). Random effects models were specified using Hedges' *g* (95% confidence interval (*CI*)) to correct

for small sample sizes in the primary studies. Hedges' *g* is interpreted similarly to Cohen's *d*: 0.2 = small effect, 0.5 = medium effect, and 0.8 = large effect (Cohen, 1988). Between-study heterogeneity of the effect sizes was tested using the I^2 statistic, which represents the proportion of observed variance in the effect size due to true heterogeneity rather than chance. According to Higgins, Thompson, Deeks & Altman (2003), an I^2 statistic of 75%, 50% and 25% indicates high, moderate, and low levels of heterogeneity, respectively. For meta-analyses including at least 10 studies publication bias was tested using a funnel plot analysis and Egger's test was performed (Higgins & Thomas, 2019; Sterne et al., 2011).

3 Results

3.1 Study selection

The study selection process is represented in Figure 1. More than 15,000 records were initially returned from database searches, and 23 additional records were identified from other sources including the reference lists of relevant reviews ($k = 16$), hand-searching journal table of contents ($k = 5$), and by scanning reference lists of included studies ($k = 2$). After removal of duplicates 12,843 records were screened for relevance based on title and abstract, with 12,582 (98.0%) excluded at this stage. Full text review against the inclusion and exclusion criteria was carried out for 261 articles. Sixty-five articles (25%) were independently reviewed by a second reviewer. There was substantial inter-rater agreement (92%; Kappa = .80, $p < .001$). Consensus was reached between the two reviewers for each case of initial disagreement.

The primary reasons for exclusion following full-text review ($k = 231$) are reported in Figure 1. Articles were most frequently excluded at this stage due to an ineligible participant population (e.g. prisoners without mental disorder), followed by lack of a suitable control or comparison condition. Methodological quality of all eligible studies was assessed independently by two reviewers using the SIGN Methodology Checklist. Three studies were excluded as the authors did not report sufficient statistical information to allow for the calculation of effect sizes (Rice, 1983; Stermac, 1986; Tomlinson & Hoaken, 2017). One study, Bernstein et al. (2012) report of preliminary findings from a Dutch trial of schema-

focussed therapy treatment, was ultimately excluded as there were insufficient studies reporting on similar outcomes for it to contribute to meta-analysis. Twenty-nine articles reporting on 28 unique studies were included in the meta-analysis.

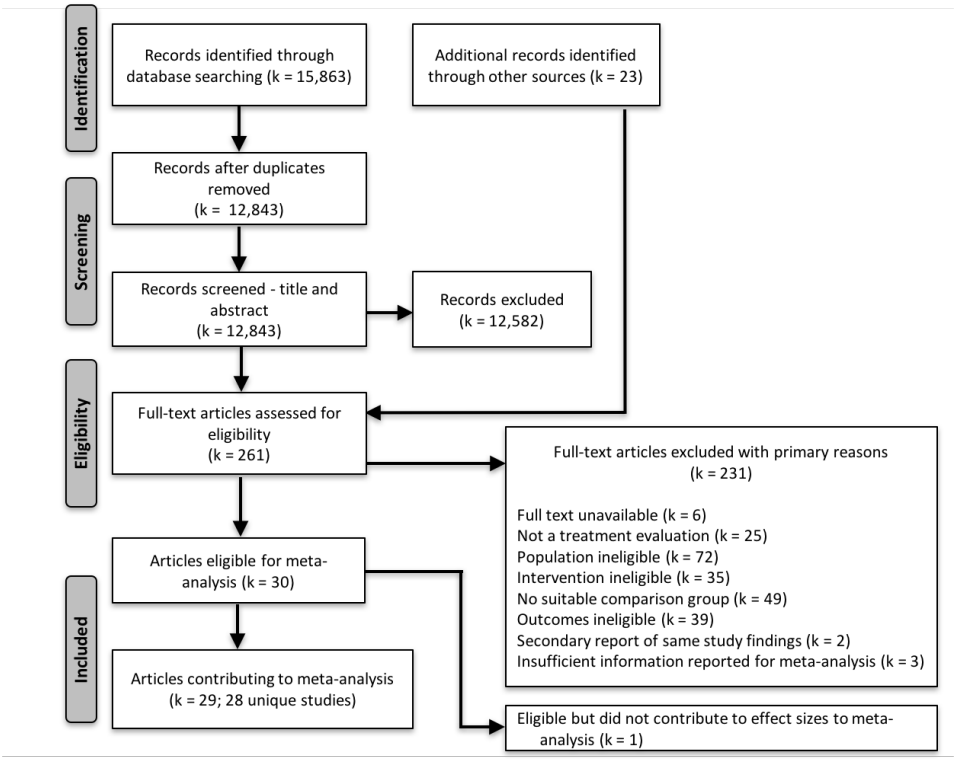


Figure 1. PRISMA flow diagram

3.2 Study characteristics

Characteristics of the included studies are presented in Table 1. Studies were published between 1985 and 2019 although the majority, 23 of the 29 papers, were published after 2010. Eleven studies followed a randomised controlled design, 12 studies were non-randomised controlled trials and five were controlled before-and-after studies. Sixteen studies were conducted in the UK, three studies in Finland, three in Canada, two each in the Netherlands and Ireland, and one study each in Italy and Australia. All studies were

carried out in forensic psychiatric hospitals except for Kingston et al. (2018) which took place in a correctional mental health treatment centre for offenders with mental illness, serving prison sentences of less than two years. The total sample size across studies was 1,422 patients, of whom 1,298 (91%) were male. Sample size ranged from 12-135, with an average size per study of 51 (M = 50.79, SD = 32.50). One study was conducted on an entirely female inpatient sample (Jotangia et al., 2015) and six studies included mixed samples. Patients' mean age was 36.54 years. In all studies, participants consented to engage in psychological treatment and in 23 studies this included written informed consent to participate in research. Studies commonly reported inclusion criteria based on diagnosis, most frequently a psychotic disorder (k = 6), schizophrenia (k = 2), or either schizophrenia or schizoaffective disorder (k = 3). Five studies limited the sample to only those with personality disorder. Three studies utilised a broader criterion of a diagnosis of severe mental illness and six studies did not state referral or inclusion criteria based on patients' diagnosis.

Table 1. Characteristics of the included studies (k = 28)

Study	Location & security level	Study design [a]	Intervention (n)	Treatment duration [b]	Treatment format	Comparison (n)	Quality [c]	MA outcomes														
								Insight	Knowledge of mental illness	Symptoms	Quality of life	Self-esteem	Coping skills	Internal locus of control	Empathy	Problem solving	Impulsivity	Ward Behaviour	Criminal and violent Attitudes	Anger and aggression	Institutional violence
Aho-Mustonen et al. (2008)	Finland High secure	NRCT	Psychoeducation (7)	1x week for 8 weeks	Group	TAU (8)	++	√	√	√												
Aho-Mustonen et al. (2011)	Finland High secure	RCT	Psychoeducation (19)	1x week for 8 weeks	Group	TAU (20)	+++	√	√	√	√	√						√				
Walker et al. (2012)	UK High secure	NRCT	Psychoeducation (28) *"Coping with Mental Illness"*	2x week for 11 weeks	Group	TAU (20)	++	√	√	√												
Walker et al. (2013)	UK High, medium, low secure Multi-site	RCT	Psychoeducation (46) *"Coping with Mental Illness"*	2x week 11 weeks	Group	Waitlist TAU (35)	++	√	√	√	√							√				
Cavezza et al. (2013)	Australia Level of security not stated	RCT	Medication adherence therapy (24)	1x week for 8 weeks	Individual	Health education intervention (24)	+	√		√												
Williams et al. (2014)	UK High secure	NRCT	CBT for schizophrenia (27)	90 min sessions 1x week 35 weeks	Group + 1:1 support	Waitlist TAU (17)	++			√												
Naughton et al. (2012)	Ireland Level of security not stated	C-B&A	Meta-cognitive training (11)	2x week for 8 weeks	Group	Waitlist TAU (8)	++			√												
Kingston et al. (2018)	Canada Correctional treatment centre; level of security not stated	RCT	R&R2 (50)	90 min sessions 2x week 7 weeks	Group	TAU (51)	++			√									√			
Kuokkanen et al. (2014)	Finland High secure	RCT	Meta-cognitive training (10)	2x week 4 weeks	Group	TAU (10)	+			√												
Mela et al. (2017)	Canada Level of security not stated	NRCT	Forgiveness strategies (36)	6 weeks	Group	Video on forgiveness (29)	+			√	√										√	

Continues on next page

Table 1. Characteristics of the included studies (k = 28)

Study	Location & security level	Study design[a]	Intervention (n)	Treatment duration[b]	Treatment format	Comparison (n)	Quality[c]	MA outcomes														
								Insight	Knowledge of mental illness	Symptoms	Quality of life	Self-esteem	Coping skills	Internal locus of control	Empathy	Problem solving	Impulsivity	Ward Behaviour	Criminal and violent Attitudes	Anger and aggression	Institutional violence
O'Reilly et al (2019)	Ireland High, medium security	RCT	Cognitive remediation training (32)	1x group + 3x individual for 14 weeks	Group + 1:1	TAU (33)	+++			√												
Tyler et al. (2018)	UK High, medium and low security	NRCT	Firesetting Intervention Programme for MDOs (63)	120 min sessions 1x week for 28 weeks	Group + 1:1	TAU (72)	+					√								√		
Donnelly & Scott (1999)	UK High secure	C-B&A	R&R (11)	120 min sessions 2x week for 5 months	Group	TAU (10)	++					√		√		√						
Yip et al. (2013)	UK High secure	NRCT	R&R MHP (30)	90 min sessions 1x week for 16 weeks	Group + 1:1 mentoring	Waitlist TAU (29)	++						√			√		√	√	√		
Zwets et al. (2016)	Netherlands High secure	RCT	Aggression Replacement Training + Psychomotor Therapy (22)	ART: 90 min sessions 1x week 35 weeks 90 min sessions 1x week 25 weeks	Group	Aggression Replacement Training + Sports sessions (15) ART – (as left) Sports – 90 min sessions, 1x week for 25 weeks	++						√						√			√	
Young et al. (2010)	UK High and Medium security	NRCT	R&R MHP (58)	90 min sessions 1x week for 16 weeks	Group + 1:1 mentoring	Waitlist TAU (12)	+						√			√		√	√			
Clarke et al. (2010)	UK Medium security	NRCT	R&R (18)	120 min sessions 36 sessions	Group	TAU (17)	+						√			√			√			
Jotangia et al. (2015)	UK Medium and low secure Multi-site	NRCT	R&R MHP (18)	90 min sessions 1x week for 16 weeks	Group + 1:1 mentoring	Waitlist TAU (20)	++							√		√		√	√	√		

Continues on next page

Table 1. Characteristics of the included studies (k = 28)

Study	Location & security level	Study design [a]	Intervention (n)	Treatment duration [b]	Treatment format	Comparison (n)	Quality [c]	MA outcomes														
								Insight	Knowledge of mental illness	Symptoms	Quality of life	Self-esteem	Coping skills	Internal locus of control	Empathy	Problem solving	Impulsivity	Ward Behaviour	Criminal and violent Attitudes	Anger and aggression	Institutional violence
Rees-Jones et al. (2012)	UK Medium and low secure Multi-site	NRCT	R&R MHP (67)	90 min sessions 1x week for 16 weeks	Group + 1:1 mentoring	Waitlist TAU (54)	++							√		√		√	√	√	
Cullen et al. (2012a & 2012b)	UK Medium secure Multi-site	RCT	R&R (44)	120 min sessions 2-3x week for 12-18 weeks	Group	TAU (40)	+								√	√			√	√	√
Lomis & Baker (1985)	Canada Medium security	RCT	Empathic communication (8)	150 min sessions 3 sessions throughout 1 day	Group	Attention Skills (8)	+								√						
Daffern et al. (2018)	UK High security	NRCT	Life Minus Violence – Enhanced (33)	120 min sessions 125 sessions for 10-12 months	Group + 1:1	Waitlist TAU (42)	+								√	√	√			√	√
Donnelly et al. (2001)	UK High secure	C-B&A	Moral reasoning (6)	90 min sessions 1x week 8 weeks	Group	TAU (6)	+								√						
Doyle et al. (2016)	UK High secure	RCT	Schema-focussed therapy (34)	1x week for > 18 months	Individual	TAU (29)	+++										√			√	
Young et al. (2013)	UK High secure	NRCT	R&R2 ADHD (16)	90 min sessions 1x week for 15 weeks	Group + 1:1 mentoring	Waitlist TAU (15)	+									√			√	√	
Bianchini et al. (2019)	Italy High secure	RCT	DBT (11)	120 min sessions 1x week for 12 months	Group + 1:1 sessions	TAU (10)	+										√				
Hornsveld & Nijman (2005)	Netherlands Level of security not stated	C-B&A	"Psychotic disorders" treatment programme (16)	12 months	Group + 1:1 sessions	TAU (16)	+											√			
Evershed et al (2003)	UK High security	C-B&A	DBT (8)	1x week for 18 months	Group +1:1 sessions	TAU (9)	+													√	√

3.3 Treatment

Nine studies evaluated variants of the Reasoning and Rehabilitation cognitive skills programme, four of which used the original full-length programme with the remainder using versions adapted for use with mentally disordered offenders. Four studies evaluated psychoeducational programmes, two studied meta-cognitive training, and two dialectical behavioural therapy (DBT). There were individual studies reporting evaluations on CBT for schizophrenia, medication adherence therapy, aggression replacement training with psychomotor therapy, schema-focussed therapy, Life Minus Violence - Enhanced, firesetting treatment, empathic communication, forgiveness skills, moral reasoning, cognitive remediation training, and a multi-component treatment programme for psychotic disorders. Treatment was most frequently delivered in groups ($k = 26$), with 12 of these studies utilising individual support or mentoring sessions in between group sessions. Only two studies (Cavezza et al., 2013; Doyle et al., 2016) evaluated treatments delivered on an individual basis. Most studies reported treatment duration of less than four months. Five studies conducted in high secure hospitals evaluated longer treatments 12-18 months. Comparison treatments were most frequently reported as treatment as usual (TAU; $k = 15$), or a waitlist condition during which patients continue TAU and would later be offered the treatment being evaluated ($k = 9$). TAU was not well described in most studies, however it was often noted by the authors to include medication, occupational therapy and social and recreational activities. Several studies reported patients in the TAU condition could and did engage in other psychological therapies during their time as control participants. Four studies used an active comparison psychological treatment that was well-matched to the experimental condition in format and duration (Cavezza et al., 2012; Lomis & Baker, 1985; Mela et al., 2016; Zwets et al., 2016).

3.4 Methodological quality

Of the 28 studies included, three were of 'high' quality (Aho-Mustonen et al., 2011; Doyle et al., 2016; O'Reilly et al., 2019) and 11 of 'acceptable' quality, however most studies ($k = 14$) were of 'low' quality. There was a tendency for older studies to have lower quality scores ($r = .31$). The most commonly failed

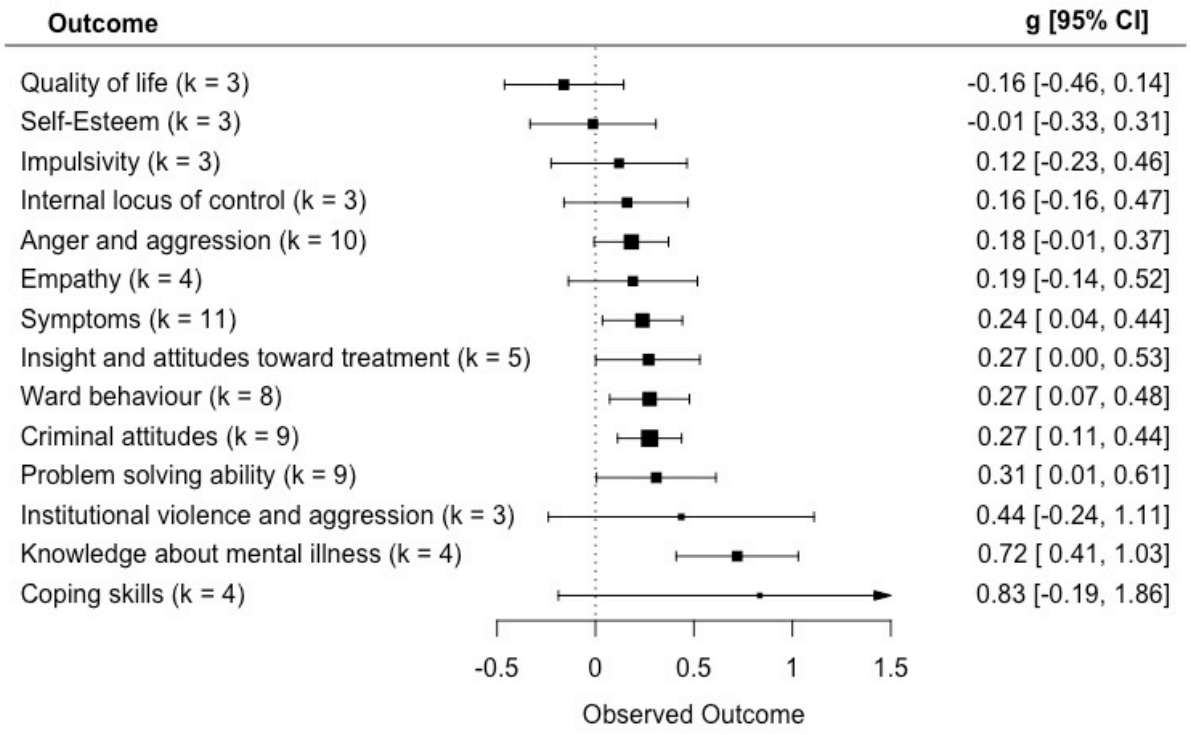
criteria included items relating to the use of an adequate randomisation method, concealment method and blinded assessments, followed by the reporting of results according to an intention-to-treat approach. Notably, several randomised controlled trials included in the review failed to meet the criteria for randomisation and concealment due to insufficient detail reported on the precise methods used for each process.

3.5 Meta-analysis

For each outcome type, a positive pooled effect size indicates improvement in patients engaged in a particular psychological treatment over and above those in the comparison condition. For most analyses there was a low level of between-study heterogeneity and therefore sensitivity analyses exploring the source of heterogeneity were not undertaken. Results are reported below for each outcome domain. A forest plot presenting the effect size for each outcome meta-analysed is shown in Figure 2. Forest plots for each individual meta-analysis are included in Appendix B.

Figure 2

Forest plot of pooled effect sizes from meta-analysis



Note: Results are presented such that positive effect sizes reflect a greater improvement in the outcome measured in the intervention being studied relative to the comparison treatment.

3.5.1 Insight into mental illness

Insight into one's mental illness was assessed as a treatment outcome in five studies, four evaluating group psychoeducational interventions compared to TAU and one evaluating individual medication adherence therapy compared to a health education intervention (Cavezza et al., 2013). Insight was assessed using various in-house designed (Aho-Mustonen et al., 2008) and well established self-report measures (Drug Attitude Inventory, Hogan, Awad & Eastwood, 1983; Patient Attitude Questionnaire, Motivation Inventory and Perception Questionnaires; Gudjonsson, Young & Yates, 2007) and semi-structured interview tools (Scale to Assess the Unawareness of Mental Disorder, Amador et al., 1994; Schedule for the Assessment

of Insight, David, 1990). Interviews were conducted by blinded assessors in Walker et al. (2013) and Aho-Mustonen et al. (2011). Treatments targeting patients' insight into their illness and attitudes towards receiving psychiatric treatment ($k = 5$, $n = 231$) had a pooled effect size of 0.27, 95% CI [0.0014, 0.53] with low between-study heterogeneity ($I^2 = 1.24\%$). The smallest effect size was observed in the study which utilised an active comparator treatment (Cavezza et al., 2013).

3.5.2 Knowledge about mental illness

Patients' knowledge about mental illness was a treatment outcome for four psychoeducational group interventions compared to TAU. In each study the outcome was assessed using a measure designed in-house by the research team. Aho-Mustonen et al. (2008; 2011) used a self-report measure called the Knowledge of Schizophrenia Scale and Walker et al. (2012; 2013) examined knowledge using the Forensic Assessment of Knowledge Questionnaire (FAKT) developed from the content of the treatment manual for the intervention under evaluation, 'Coping with Mental Illness.' The FAKT was administered by an assessor blind to treatment condition in Walker et al. (2013). There was a medium pooled effect in favour of psychoeducation in patients' knowledge of mental illness relative to TAU ($k = 4$, $n = 179$, $g = 0.72$ [0.41, 1.03], $I^2 = 0\%$).

3.5.3 Symptoms

Psychiatric symptoms were measured in eleven studies. A range of therapies are represented in these studies, including psychoeducation, meta-cognitive training, cognitive remediation training, medication adherence therapy, CBT for schizophrenia, and Reasoning & Rehabilitation for Mentally Disordered Offenders (R&R2M). Except for CBT for schizophrenia evaluated in Williams et al. (2014) which was delivered over nine months, these interventions were brief, lasting less than three months. Symptoms were measured by various self-report (Beck Depression Inventory, Beck, Steer, & Brown, 1996; Depression

Anxiety Stress Scale, Lovibond & Lovibond, 1995) and clinician-rated interview scales (Brief Psychiatric Rating Scale, Overall & Gorham, 1962; Calgary Depression Scale for Schizophrenia, Addington, Addington, & Maticka-Tyndale, 1993; Positive and Negative Symptoms Scale, Kay et al., 1987; Psychotic Symptom Rating Scales, Haddock et al., 1999; Scale for the Assessment of Negative Symptoms, Andreasen, 1983; Scale for the Assessment of Positive Symptoms, Andreasen, 1984). Psychological therapies resulted in a small improvement in patients' symptoms (k = 11, n = 521, g = 0.24 [0.04, 0.44], I^2 = 21.83%). Funnel plot analysis (Appendix C) did not indicate publication bias and Egger's test was non-significant ($t(9)$ = -0.27, p = .79).

3.5.4 Quality of Life

Forensic patients' quality of life was measured by three studies; two psychoeducational interventions (Aho-Mustonen et al., 2011; Walker et al., 2013) and an evaluation of a brief forgiveness skills training (Mela et al., 2016). Each study used a different patient self-report questionnaire to measure this, including Sintonen's (2001) 15D instrument of health related quality of life, Diener's Satisfaction with Life Scale (Pavot & Diener, 1993) and the Schizophrenia Quality of Life Scale Revision 4 (Martin & Allan, 2007). The effect of psychological therapy on quality of life was not significant over and above the comparator treatment (k = 3, n = 171, g = -0.16 [-0.46, 0.14], I^2 = 0%).

3.5.5 Self-Esteem

Three studies measured the effects of treatment on self-esteem. The treatments were varied and included Reasoning & Rehabilitation (Donnelly & Scott, 1999), psychoeducation (Aho-Mustonen et al., 2011) and a firesetting treatment (Tyler et al., 2018). The Rosenberg Self Esteem Scale (1965) and Culture-Free Self-Esteem Inventory (Battle, 1992) were used. Meta-analysis found no effect of psychological treatment on patients' self-esteem over the comparator treatment (k = 3, n = 152, g = -0.01 [-0.33, 0.31], I^2 = 0%).

3.5.6 Coping skills

Coping skills was reported in four studies, three of which were evaluations of Reasoning and Rehabilitation (Clarke et al., 2010; Yip et al., 2013; Young et al., 2010). Zwets et al. (2016) also studied the effects of Aggression Replacement Training with Psychomotor Therapy on coping skills. Self-report scales were used, including The Coping Responses Inventory (Moos, 1993), Utrecht Coping Scale (Schreurs et al., 1993), and the Ways of Coping Scale (Lazarus & Folkman,1984). The overall effect of treatment was not significant (k = 4, n = 126, g = 0.83 [-0.19, 1.86], I^2 = 85%). There was high heterogeneity observed, due to two studies having a near-zero effect size (Young et al., 2010; Zwets et al., 2016) and two with very large positive effects favouring the Reasoning & Rehabilitation programmes being evaluated (Yip et al., 2013; Clarke et al., 2010). Though there are too few studies in this meta-analysis to support a sensitivity analysis, consideration was given to whether differences in treatment type, study design, measurements used, and study quality could account for between-study variation of treatment effect. No clear differences emerged that might account for the very different effect sizes observed.

3.5.7 Locus of control

Locus of control, or the perceived control over one's own life and problems, was assessed in three studies. Each study evaluated a Reasoning and Rehabilitation programme relative to TAU. In Donnelly & Scott (1999) the original R&R programme was used, delivered in group sessions twice weekly over 5-months. In Jotangia et al (2015) and Rees-Jones et al. (2012) the revised R&R2M was delivered over a 16-week period, with individual mentoring between group sessions. Each study used the Locus of Control Scale by Nowicki & Duke (1974). Meta-analysis found that R&R was not effective in increasing patients' internal locus of control over and above TAU (k = 3, n = 180, g = 0.16 [-0.16, 0.47], I^2 = 5.97%).

3.5.8 Empathy

Four studies evaluated the effects of treatment on patients' self-reported empathy. Treatments included ranged from R&R (Cullen et al., 2012); LMV-E (Daffern et al., 2018), moral reasoning (Donnelly et al., 2001) and microtraining on empathic communication skills (Lomis & Baker, 1985). Various measures were employed, including the Interpersonal Reactivity Index (IRI; Davis,), an empathy subscale from the Impulsiveness Questionnaire, and three different empathy measures each measuring a different facet of empathy (state, trait, and cognitive empathy) in Lomis & Baker (1985) evaluation of empathic skills training. The largest treatment effect was observed in Lomis & Baker (1985). The overall treatment effect was not significant ($k = 4$, $n = 147$, $g = 0.19$ [-0.14, .52], $I^2 = 0\%$).

3.5.9 Problem-solving

Problem-solving ability was assessed in nine studies. All but one of these (Daffern et al., 2018) evaluated variants of R&R compared to treatment as usual, with Daffern and colleagues evaluating Life Minus Violence- Enhanced (LMVE-E). Like R&R, LMV-E aims to develop patients' cognitive and interpersonal skills in order to avoid situations resulting in violent or aggressive behaviour. In every study except for Donnelly & Scott (1999), which used the Means-Ends Problem Solving Procedure (MEPS; Platt & Spivack, 1975), the D'Zurilla et al (2002) Social Problem Solving Inventory Revised Short Form was used to assess social problem-solving ability. Treatment yielded a small improvement in patients' problem-solving ability ($k = 9$, $n = 449$, $g = 0.31$ [0.01, 0.61]. Moderate between-study heterogeneity was observed ($I^2 = 56.86\%$).

3.5.10 Impulsivity

Three studies reported the effects of treatment on patients' self-reported impulsivity. Treatments assessed include LMV-E (Daffern et al., 2018), DBT (Bianchini et al., 2019) and individual schema-focussed therapy (Doyle et al., 2016). The three interventions were between 12 and18 months in duration. There were some

similarities in the population targeted by the three studies. The samples in Daffern et al. (2018) and Bianchini et al. (2019) were recruited based on a history of interpersonal violence. Bianchini et al. (2019) and Doyle et al. (2016) sample comprised of men with personality disorder, while 37% of participants in Daffern et al. (2018) had a diagnosis of PD (either primary or comorbid). In each study the Barratt Impulsiveness Scale (Barratt, 1994) was used to measure impulsivity. The treatment effect was not significant (k = 3, n = 141, g = 0.12 [-0.23, 0.46], I^2 = 0%).

3.5.11 Ward behaviour

Ward behaviour was reported in eight studies including two group psychoeducation interventions (Aho-Mustonen et al., 2011; Walker et al., 2013), R&R2M evaluations, Zwets et al (2016) trial of psychomotor therapy as an adjunct component of Aggression Replacement Training, and Hornsveld & Nijman's evaluation of a psychotic disorders treatment programme. In each study, ward behaviour was rated by informants, typically ward-based nursing staff who were not blind to participants' allocation except for Walker et al (2013) where the rating was conducted by a researcher blind to patients' allocated treatment. The Disruptive Behaviour and Social Problem Solving Scale (DBSP; Young, Gudjonsson, Ball & Lam, 2003), Nurses' Observation Scale for Inpatient Evaluation (NOSIE-30; Honigfeld, Roderic & Klett, 1966), Behaviour Status Index (BEST Index; Woods, Reed & Robinson, 1999), the Observation Scale for Aggressive Behaviour (OSAB; Hornsveld, Nijman, Hollin & Kraaimaat, 2007), and a Dutch instrument called the Meijers Insitute Observation Scale (Brand et al., 1998) were used. Positive effect sizes represent more frequent pro-social behaviour and/or less antisocial or otherwise problematic behaviour reported after treatment. Meta-analysis found a small effect of psychological treatment on ward-based social behaviour relative to the comparator condition (k = 8, n = 392, g = 0.27 [0.07, 0.48], I^2 = 0%).

3.5.12 Criminal attitudes

Nine studies measured the effect of treatment on self-reported criminal attitudes; eight studies evaluating variants of R&R, as well as Tyler et al's (2018) evaluation of firesetting treatment programme. In each case the comparator was TAU or waitlist TAU. Studies used various self-report measures including the Maudsley Violence Questionnaire (Walker, 2005) the Crime Pics II (Frude et al., 1994), the Measure of Criminal Attitudes and Associates (Mills & Kroner, 1999), and the Criminal Sentiments Scale-Modified (Simourd, 1997). There was a small effect for a reduction in pro-criminal and violent attitudes due to psychological treatment relative to treatment as usual (k = 9, n = 561, g = 0.27 [0.11, 0.44], I^2 = 0%).

3.5.13 Anger and aggression

Anger was reported as a treatment outcome in ten studies, five evaluations of R&R2M, an individual schema-focussed therapy for patients with personality disorder (Doyle et a., 2016), LMV-E (Daffern et al., 2018), DBT (Evershed et al., 2003), a trial of Aggression Replacement Training with Psychomotor Therapy (ART + PMT) compared to ART with sports sessions (Zwets et al., 2016), and a forgiveness skills intervention (Mela et al., 2016). Each study used a self-report measure, either the Novaco Anger Scale (NAS; Novaco, 1994), the Novaco Anger Scale and Provocation Inventory (NAS-PI; Novaco, 2003) or the Spielberger State-Trait Anger Expression Inventory-2 (STAXI-2; Spielberger, 1999) to measure anger. In addition to the NAS, Zwets et al (2016) included the Aggression Questionnaire Short Form (Bryant & Smith, 2001). Psychological intervention was not effective in reducing patients' self-reported anger or aggression over and above the comparator treatment (k = 10, n = 528, g = 0.18 [-0.01, 0.37], I^2 = 19.29%). Funnel plot analysis (Appendix C) did not indicate publication bias and Egger's test was non-significant ($t(8)$ = 0.99, p = .35).

3.5.14 Institutional violence and aggression

The effects of treatment on the frequency of recorded incidents of violence or aggression in an inpatient setting was reported by three studies. The treatments evaluated in each study (R&R; Cullen et al., 2012b; LMV-E, Daffern et al., 2018; DBT, Evershed et al., 2003) had a core aim of reducing these behaviours, and individuals with a history of violent offending or behaviour were recruited for therapy. Evidence of violence or aggression were recorded from patients' notes and the number of incidents was summed from a period before treatment began and during a follow up period after treatment had completed (except for Daffern et al., 2018 for which the latter timepoint reflected the number of incidents during the 12-month LMV-E treatment). Incidents were also recorded for comparison participants during a similar timeframe. The effect of treatment on institutional violence and aggression was not significant ($k = 3$, $n = 141$, $g = 0.44$ [-0.24, 1.11]). There was moderate between-study heterogeneity observed ($I^2 = 67.79\%$). *T*he reduction in violent behaviour from DBT treatment reported by Evershed et al (2003) was substantially larger than the effect size reported by the other two studies.

4 Discussion

Despite widespread use of psychological and psychosocial therapies with forensic mental health patients, there is a lack of empirical evidence supporting the use of such treatments (Barnao & Ward, 2015; Mallion, Tyler & Miles, 2019). Determining the effectiveness of interventions delivered is necessary to embedding the principle of evidence-based practice in forensic mental health services. This paper reported a systematic review and meta-analysis of controlled evaluations of psychological interventions for forensic mental health inpatients based on 28 studies involving 1,422 individuals. To the authors' knowledge, this is the first meta-analytic review which focuses on psychological and psychosocial interventions delivered in forensic mental health hospitals.

Studies of psychological interventions in forensic hospitals which adopt a controlled design are becoming more frequent, evidenced by the fact that most studies were published after 2010, despite those

from 1980 onward being eligible. A majority of studies (k = 16) were conducted in the United Kingdom and variants of Reasoning and Rehabilitation cognitive skills programmes appear to be the most frequently studied psychological intervention in this population and treatment setting, followed by studies evaluating psychoeducational interventions. With few exceptions, studies adopted well-established outcome measures known to have robust psychometric properties in other clinical populations. Notably, the psychoeducation studies employed in-house designed outcome measures to test participants' knowledge and understanding of their mental illness. This outcome yielded the largest significant treatment effect size (g = 0.72 [0.41, 1.03]), however caution should be used not to over generalise this finding as it may be partially inflated by the authors having designed an outcome measure to match the content taught in the specific intervention (so called "teaching to the test"). Studies evaluating Reasoning and Rehabilitation programmes adopted similar assessment batteries which aided greatly in the comparison of results with other R&R evaluations. It should be noted that despite the overwhelming use of standardised outcome measures by studies, there are few outcome measures for which the validity and psychometric properties in forensic populations has been assessed and this remains an understudied area (Chamber et al., 2009) with great potential to aid advancement of the evidence base for forensic mental health treatment.

Pooled effect sizes for fourteen outcomes were calculated. No effect of treatment was observed for eight outcome domains including patients' quality of life, self-esteem, impulsivity, anger, locus of control, empathy, coping skills, and recorded incidents of violence and aggression. Small treatment effects were observed for insight, psychiatric symptoms, problem-solving ability, criminal attitudes, and observer-rated ward behaviour. One medium effect size was found, for group psychoeducation on patients' knowledge of mental disorder. For most analyses, heterogeneity calculated by the I^2 statistic was low, suggesting consistency between studies reporting on similar outcome domains (Higgins et al., 2003). Treatment effects found in this review are smaller than those for mental health outcomes in prisoners reported by the Yoon, Slade & Fazel (2017) review. Yoon et al. (2017) reported an overall pooled effect size (ES) of 0.50 (95% *CI* 0.34-0.66) for psychological treatments in randomised trials involving prisoners with mental health problems. Effect sizes for specific outcomes such as depression (*ES* = 0.60, 95% *CI* 0.38-0.83), anxiety

(*ES* = 0.56, 95% *CI* 0.31-0.82), and hostility/ anger (*ES* = 0.42, 95% *CI* 0.13-0.71) exceed all effects found in this review with the exception of forensic patients' knowledge about mental illness. However, the variance may be due in large part to prevalence of different control conditions by the primary studies in the two reviews. In Yoon et al. (2017), 16% of studies utilised a no treatment control condition, 41% used a waitlist control (during which participants did not access any treatment), and 43% used TAU or another psychological therapy as the comparison condition (during which prisoners received some form of treatment). In contrast, all 28 studies in the present review fall within the latter category (in nine studies the TAU participants were on a waitlist for the intervention being studied). As the effect sizes from the present meta-analysis quantify the intervention effect over and above that of other co-occurring treatments they would be expected to be smaller than those derived from no-treatment comparisons. Every included study took place in a psychiatric treatment facility where participants were involuntarily detained for treatment; it is difficult to conceive that a study could ethically be carried out in such an environment which adopts a comparison condition where participants receive no treatment at all. This is a critical distinction from intervention trials taking place in prison or custodial settings. The present findings are broadly consistent with the small effect sizes reported by Yoon et al. (2017) for studies which adopted a comparison condition during which participants received some form of psychological therapy (*ES* = 0.21 (95% 0.01-0.41)). Considered together, the findings of the present review and Yoon et al (2017) indicate that psychological treatments delivered to mentally disordered offenders, whether in hospital or prison, may offer only modest benefit over and above concurrent treatments.

Psychological interventions delivered in forensic mental health care can be extremely resource-intensive, requiring a team of trained and experienced practitioners (often multi-disciplinary) delivering group treatment up to several times per week to a small group of patients. In many cases, such intensive group psychological treatment is supplemented with individual support or mentoring between group sessions (as was the case for 12 studies in this review). Individual treatment can draw on significant staff resources. Of course, resource-intensive, highly specialist treatment may be fully warranted to the extent that treatment leads to improvement in mental health or criminogenic needs which are not targeted by other

less intensive interventions, or the improvements exceed those likely to be achieved by other interventions, or even that the intensive therapy motivates and engages patients who would not engage in other therapy. It seems likely there will be a continued role for the use of highly specialist and resource-intensive treatments with forensic mental health patients, though it remains to be examined whether such treatments offered alongside brief, low intensity group interventions as part of a stepped care model of service delivery (Bower & Gilbody, 2005) can be both an effective and efficient method of operating psychological services in forensic settings.

Assessment of study methodological quality using the SIGN checklist indicated half of the studies (14/28) in this review were of low quality. The quality criteria frequently related to decisions in the study design, data analysis, and reporting stages. There was insufficient information reported on randomisation, allocation concealment and blinding in several of the included RCTs, indicating a need to raise awareness of and encourage adherence to the Consolidated Standards of Reporting Trials (CONSORT; Schulz et al., 2010) guidance and its subsequent extension to trials of social and psychological interventions (CONSORT-SPI 2018; Grant et al., 2018), which require complete reporting of these study design elements. Few studies used the intention-to-treat principle in analysing data and reporting results, opting instead to use a per-protocol or complete-case analysis, which is known to yield biased results (Armijo-Olivo, Warren, & Magee, 2009). This is problematic for the following two reasons. Firstly, psychological therapy in forensic mental health patients is wrought with poor attendance and attrition, with attrition more likely for certain subgroups of patients. For example, certain characteristics, including recent violence behaviour, antisocial traits and psychopathy, are associated with forensic patients who discontinue cognitive skills therapy compared to those who complete it (Cullen et al., 2011). Secondly, the defining criteria for a 'completer' in the therapies studied in this review often appeared to be arbitrarily decided based on a proportion of the number of sessions attended. Readers should be cautious of treatment evaluations using only a per-protocol analysis in the absence of a sound justification by the authors for their operationalisation of 'treatment completion.'

This review reveals a perhaps surprising lack of studies evaluating therapies which are commonly delivered in forensic mental health hospitals, for example substance misuse or sexual offending behaviour treatment programmes. From a UK perspective, according to the 29 forensic mental health units surveyed by Mallion et al (2019), 25 had provided group treatment for substance misuse, and 17 had delivered sexual offending treatment within the preceding 5 years. The Mallion et al. (2019) review clearly demonstrates the disconnect between the therapies delivered in practice and the therapies arising in the published research literature. With the substantial and sustained growth in this area of research over the past 10 years it is hoped this gap will close in time.

4.1 Limitations

There are limitations to this review. The limited number of studies in each meta-analysis precluded our ability to test of publication bias for treatment effects on 12 of 14 outcomes. However, the search strategy was designed to identify relevant grey literature, though ultimately all included studies had been published. Encouragingly, there was no evidence for publication bias in the studies reporting on the effects of treatment on patients' symptoms, and on self-reported anger or aggression. In addition to bias toward published papers reporting significant results (rather than null), researchers and clinicians should be wary of the existence of unpublished studies which may have in fact found significant effects in the unintended direction, evidence of adverse effects of treatment (Duggan et al., 2014; McIntosh et al., 2019). Indeed, the MacInnes & Masino (2019) review highlighted several studies which found the psychological intervention led to statistically significant negative effects on outcomes of violence or risk of future violence as well as a measure of ward atmosphere. Several studies in the present review also reported a negative treatment effect relative to the comparison condition (Cavezza et al 2013; Zwets et al 2016).

Quality assessment revealed limitations in the design and reporting of the included studies which cause uncertainty in the estimated treatment effect. Studies with high risk of bias leave open greater potential that the observed outcome differences between groups could have been due to factors other than

the treatment being evaluated. It is therefore possible that the pooled treatment effect sizes reported by this review (most of which were small or not significant) in fact overestimate the true effect of these treatments.

Women comprised just 9% of the total sample in this review, which limits generalisability of the findings to female forensic patients. This gender split is representative of the small minority of women who access forensic mental health inpatient services and the broader difficulties in identifying mental health interventions supported by evidence as efficacious for female forensic patients (Tolland et al., 2019). To overcome this problem, researchers evaluating psychological therapies in forensic samples which include women should report subgroup results for the female sample or make anonymised individual patient datasets of these evaluations available to researchers undertaking meta-analyses.

4.2 Conclusions

This review found that psychological interventions delivered to forensic mental health inpatients show small effects over and above the comparator treatment for insight, symptoms, problem-solving ability, reduced pro-criminal attitudes and improved ward behaviour. A medium effect was found for treatment increasing patients' knowledge of their mental illness. Results showed that psychological treatment had no benefit over the comparator condition on many relevant domains including impulsivity, empathy, coping skills, anger or inpatient violence. As this is a rapidly growing area of research and service development in forensic healthcare, future reviews on the effectiveness are likely to follow. The focus of such future research should be to identify the treatments that are most effective, as well as determine the patient characteristics more closely associated with treatment benefit. Given the modest treatment effects observed, these findings would also warrant an examination of the cost-effectiveness of psychological therapies for forensic mental health patients.

References

* indicates study was included in meta-analysis

Addington, D., Addington, J., & Maticka-Tyndale, E. (1993). Assessing depression in schizophrenia: the Calgary Depression Scale. *The British Journal of Psychiatry*, *163*(S22), 39-44. doi: 10.1192/S0007125000292581

*Aho-Mustonen, K., Miettinen, R., Koivisto, H., Timonen, T., & Räty, H. (2008). Group psychoeducation for forensic and dangerous non-forensic long-term patients with schizophrenia: A pilot study. *The European Journal of Psychiatry*, *22*(2), 84-92.

*Aho- Mustonen, K., Tiihonen, J., Repo- Tiihonen, E., Ryynänen, O. P., Miettinen, R., & Räty, H. (2011). Group psychoeducation for long- term offender patients with schizophrenia: An exploratory randomised controlled trial. *Criminal Behaviour and Mental Health*, *21*(3), 163-176. doi: 10.1002/cbm.788

Amador, X. F., Flaum, M., Andreasen, N. C., Strauss, D. H., Yale, S. A., Clark, S. C., & Gorman, J. M. (1994). Awareness of illness in schizophrenia and schizoaffective and mood disorders. *Archives of General Psychiatry*, *51*(10), 826-836. doi: 10.1001/archpsyc.1994.03950100074007

Andreasen, N.C. (1983). The scale for the assessment of negative symptoms (SANS). University of Iowa, Iowa City.

Andreasen, N.C. (1984). The scale for the assessment of positive symptoms (SAPS). University of Iowa, Iowa City.

Armijo-Olivo, S., Warren, S., & Magee, D. (2009). Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. *Physical Therapy Reviews*, *14*(1), 36-49. doi: 10.1179/174328809X405928

Battle, J. (1992). *Culture-free self-esteem inventories.* Austin, TX: Pro Education.

Barnao, M., & Ward, T. (2015). Sailing uncharted seas without a compass: A review of interventions in forensic mental health. *Aggression and Violent Behavior*, *22*, 77-86. doi: 10.1016/j.avb.2015.04.009

Barratt, E.S. (1994). Impulsiveness and aggression. In J. Monahan & H. J. Steadman (Eds.), *Violence and mental disorder: Developments in risk assessment* (pp. 61-80). Chicago, IL: University of Chicago Press.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Manual for the Beck Depression Inventory-II. 1996. *San Antonio, TX: Psychological Corporation*.

Bernstein, D. P., Nijman, H. L., Karos, K., Keulen-de Vos, M., de Vogel, V., & Lucker, T. P. (2012). Schema therapy for forensic patients with personality disorders: Design and preliminary findings of a multicenter randomized clinical trial in the Netherlands. *International Journal of Forensic Mental Health*, *11*(4), 312-324. doi: 10.1080/14999013.2012.746757

*Bianchini, V., Cofini, V., Curto, M., Lagrotteria, B., Manzi, A., Navari, S., ... & Silvestrini, C. (2019). Dialectical behaviour therapy (DBT) for forensic psychiatric patients: An Italian pilot study. *Criminal Behaviour and Mental Health*, *29*(2), 122-130. doi: 10.1002/cbm.2102

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: access, effectiveness and efficiency: narrative literature review. *The British Journal of Psychiatry*, *186*(1), 11-17. doi: 10.1192/bjp.186.1.11

Brand, E. F. J. M., Diks, G. J. M., van Emmerik, J. L., & Raes, B. C. M. (1998). *Diagnostiek en onderzoek [Assessment and Research]. Utrecht, NL: Dr. F. S. Meijers Instituut.*

Bryant, F. B., & Smith, B. D. (2001). Refining the architecture of aggression: A measurement model for the Buss–Perry Aggression Questionnaire. *Journal of Research in Personality*, *35*(2), 138-167. doi:10.1006/jrpe.2000.2302

*Cavezza, C., Aurora, M., & Ogloff, J. R. (2013). The effects of an adherence therapy approach in a secure forensic hospital: a randomised controlled trial. *Journal of Forensic Psychiatry & Psychology*, *24*(4), 458-478. doi: 10.1080/14789949.2013.806568

Chambers, J. C., Yiend, J., Barrett, B., Burns, T., Doll, H., Fazel, S., . . . Plugge, E. (2009). Outcome measures used in forensic mental health research: a structured review. *Criminal Behaviour and Mental Health, 19*(1), 9-27. doi: 10.1002/cbm.724

*Clarke, A. Y., Cullen, A. E., Walwyn, R., & Fahy, T. (2010). A quasi-experimental pilot study of the Reasoning and Rehabilitation programme with mentally disordered offenders. *The Journal of Forensic Psychiatry & Psychology*, *21*(4), 490-500. doi: 10.1080/14789940903236391

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.) Hillsdale, New Jersey: Lawrence Earlbaum Associates.

*Cullen, A. E., Clarke, A. Y., Kuipers, E., Hodgins, S., Dean, K., & Fahy, T. (2012). A multisite randomized trial of a cognitive skills program for male mentally disordered offenders: Violence and antisocial behavior outcomes. *Journal of Consulting and Clinical Psychology*, *80*(6), 1114-1120. doi: 10.1037/a0030291

*Cullen, A. E., Clarke, A. Y., Kuipers, E., Hodgins, S., Dean, K., & Fahy, T. (2012). A multi-site randomized controlled trial of a cognitive skills programme for male mentally disordered offenders: social–cognitive outcomes. *Psychological Medicine*, *42*(3), 557-569. doi: 10.1017/S0033291711001553

Cullen, A. E., Soria, C., Clarke, A. Y., Dean, K., & Fahy, T. (2011). Factors predicting dropout from the Reasoning and Rehabilitation Program with mentally disordered offenders. *Criminal Justice and Behaviour, 38*(3), 217-230. doi: 10.1177/0093854810393659

*Daffern, M., Simpson, K., Ainslie, H., & Chu, S. (2018). The impact of an intensive inpatient violent offender treatment programme on intermediary treatment targets, violence risk and aggressive behaviour in a sample of mentally disordered offenders. *The Journal of Forensic Psychiatry & Psychology*, *29*(2), 163-188. doi: 10.1080/14789949.2017.1352014

David, A. S. (1990). Insight and psychosis. *The British Journal of Psychiatry*, *156*(6), 798-808. doi: 10.1192/bjp.156.6.798

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. JSAS *Catalog of Selected Documents in Psychology*, *10*, 85.

*Donnelly, J. P., & Scott, M. F. (1999). Evaluation of an offending behaviour programme with a mentally disordered offender population. *British Journal of Forensic Practice*, *1*(4), 25-32.

*Donnelly, J., Williamson, L., & Duncan, E. (2001). Facilitating moral reasoning. *Forensic Update*, *67*, 5-11.

*Doyle, M., Tarrier, N., Shaw, J., Dunn, G., & Dolan, M. (2016). Exploratory trial of schema-focussed therapy in a forensic personality disordered population. *The Journal of Forensic Psychiatry & Psychology*, *27*(2), 232-247. doi: 10.1080/14789949.2015.1107119

Duggan, C., Parry, G., McMurran, M., Davidson, K., & Dennis, J. (2014). The recording of adverse events from psychological treatments in clinical trials: evidence from a review of NIHR-funded trials. *Trials*, *15*(1), 335. doi: 10.1186/1745-6215-15-335

Dumont, M., Thériault, J., Briand, C., Dumais, A., & Potvin, S. (2018). Psychosocial approaches for individuals with schizophrenia in correctional and forensic psychiatric settings: a rapid review. *Journal of Forensic Practice, 20*(3), 152-166. doi: 10.1108/JFP-11-2017-0049

Duncan, E. A., Nicol, M. M., Ager, A., & Dalgleish, L. (2006). A systematic review of structured group interventions with mentally disordered offenders. *Criminal Behaviour and Mental Health*, *16*(4), 217-241. doi: 10.1002/cbm.631

Durcan, G., Hoare, T., & Cumming, I. (2011). Unlocking Pathways to Secure Mental HealthCare. Centre for Mental Health. Accessed on www.centreformentalhealth.org.uk on 30 January 2020.

D'Zurilla, T. J., Nezu, A. M., Maydeu-Olivare, A., Olivares, A., & D'Zurilla, T. (2002). Social problem-solving inventory–Revised (SPSI–R).

*Evershed, S., Tennant, A., Boomer, D., Rees, A., Barkham, M., & Watson, A. (2003). Practice- based outcomes of dialectical behaviour therapy (DBT) targeting anger and violence, with male forensic patients: A pragmatic and non- contemporaneous comparison. *Criminal Behaviour and Mental Health*, *13*(3), 198-213. doi: 10.1002/cbm.542

Eysenck, H., J. & Eysenck, S. B. G. (1991). *Manual of Eysenck Personality Scales (EPS Adult)*. London Hodder and Stoughton.

Frude, N., Honess, T., & Maguire, M. (1994). *Crime-Pics II: Manual*. Michael & Associates.

Gudjonsson, G. H., Young, S., & Yates, M. (2007). Motivating mentally disordered offenders to change: instruments for measuring patients' perception and motivation. *The Journal of Forensic Psychiatry & Psychology*, *18*(1), 74-89. doi: 10.1080/14789940601063261

Grant, S., Mayo-Wilson, E., Montgomery, P., Macdonald, G., Michie, S., Hopewell, S., & Moher, D. (2018). CONSORT-SPI 2018 Explanation and Elaboration: guidance for reporting social and psychological intervention trials. *Trials*, *19*(1), 406.

Haddock, G., McCarron, J., Tarrier, N., & Faragher, E. B. (1999). Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS). *Psychological Medicine*, *29*(4), 879-889. doi: doi: 10.1007/s001270050141

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. version 6.0 (updated July 2019). Available from www.training.cochrane.org/handbook.

Higgins, J. P., Thompson, S.G., Deeks, J. J., Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557-560. doi: 10.1136/bmj.327.7414.557

Hofmann, S., A, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy Research, 36*(5), 427-440. doi: 10.1007/s10608-012-9476-1

Hogan, T. P., Awad, A. G., & Eastwood, R. (1983). A self-report scale predictive of drug compliance in schizophrenics: reliability and discriminative validity. *Psychological Medicine*, *13*(1), 177-183. doi: 10.1017/S0033291700050182. doi:

Honigfeld, G., Gillis, R. D., & Klett, C. J. (1966). NOSIE-30: A treatment-sensitive ward behavior scale. *Psychological Reports*, *19*(1), 180-182. doi: 10.2466/pr0.1966.19.1.180

*Hornsveld, R. H., & Nijman, H. L. (2005). Evaluation of a cognitive-behavioral program for chronically psychotic forensic inpatients. *International Journal of Law and Psychiatry*, *28*(3), 246-254. doi: 10.1016/j.ijlp.2004.09.004

Hornsveld, R. H., Nijman, H. L., Hollin, C. R., & Kraaimaat, F. W. (2007). Development of the Observation Scale for Aggressive Behavior (OSAB) for Dutch forensic psychiatric inpatients with an antisocial personality disorder. *International Journal of Law and Psychiatry*, *30*(6), 480-491.

*Jotangia, A., Rees-Jones, A., Gudjonsson, G. H., & Young, S. (2015). A multi-site controlled trial of the R&R2MHP cognitive skills program for mentally disordered female offenders. *International Journal of Offender Therapy and Comparative Criminology*, *59*(5), 539-559. doi: 10.1177/0306624X13512092

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, *13*(2), 261-276. doi: 10.1093/schbul/13.2.261

*Kingston, D. A., Olver, M. E., McDonald, J., & Cameron, C. (2018). A randomised controlled trial of a cognitive skills programme for offenders with mental illness. *Criminal Behaviour and Mental Health*, *28*(4), 369-382. doi: 10.1002/cbm.2077

*Kuokkanen, R., Lappalainen, R., Repo- Tiihonen, E., & Tiihonen, J. (2014). Metacognitive group training for forensic and dangerous non- forensic patients with schizophrenia: A randomised controlled feasibility trial. *Criminal Behaviour and Mental Health*, *24*(5), 345-357. doi: 10.1002/cbm.1905

Lazarus, R., & Folkman, R. S. (1984). *Stress, appraisal and coping*. New York: Springer.

*Lomis, M. J., & Baker, L. L. (1985). Microtraining of forensic psychiatric patients for empathic counseling skills. *Journal of Counseling Psychology*, *32*(1), 84-93. doi: 10.1037/0022-0167.32.1.84

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335-343. doi: 10.1016/0005-7967(94)00075-U

MacInnes, D., & Masino, S. (2019). Psychological and psychosocial interventions offered to forensic mental health inpatients: a systematic review. *BMJ Open*, *9*(3), e024351. doi: 10.1136/bmjopen-2018-024351

Mallion, J.S., Tyler, N., & Miles, H.L. (in press). What is the evidence for offence-specific group treatment programs for forensic patients? *International Journal of Forensic Mental Health.* doi: 10.1080/14999013.2019.1648344

Martin, C. R., & Allan, R. (2007). Factor structure of the schizophrenia quality of life scale revision 4 (SQLS-R4). *Psychology, Health & Medicine*, *12*(2), 126-134. doi: 10.1080/13548500500407383

Martin, M. S., Dorken, S. K., Wamboldt, A. D., & Wootten, S. E. (2012). Stopping the revolving door: A meta-analysis on the effectiveness of interventions for criminally involved individuals with major mental disorders. *Law and Human Behavior*, *36*(1), 1-12. doi: 10.1037/h0093963

McIntosh, L. G., McMurran, M., Taylor, P. J., & Thomson, L. D. (2019). Gaps in measures of adverse outcomes relating to psychological interventions. *Criminal Behaviour and Mental Health*, *29*(1), 1-6. doi: 10.1002/cbm.2100

*Mela, M., Baetz, M., Marcoux, G., Delury, D., Cooper, B., & Sajobi, T. T. (2017). The influence of a learning to forgive programme on negative affect among mentally disordered offenders. *Criminal Behaviour and Mental Health*, *27*(2), 162-175. doi: 10.1002/cbm.1991

Mills, J. F., & Kroner, D. G. (1999). Measures of criminal attitudes and associates: User guide. *Unpublished instrument and user guide*.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264-269. doi: 10.7326/0003-4819-151-4-200908180-00135

Moos, R. (1993). *Coping responses inventory – adult form professional manual.* Odessa, FL: PAR.

Morgan, R. D., Flora, D. B., Kroner, D. G., Mills, J. F., Varghese, F., & Steffan, J. S. (2012). Treating offenders with mental illness: A research synthesis. *Law and Human Behavior*, *36*(1), 37-50. doi: 10.1037/h0093964

*Naughton, M., Nulty, A., Abidin, Z., Davoren, M., O'Dwyer, S., & Kennedy, H. G. (2012). Effects of group metacognitive training (MCT) on mental capacity and functioning in patients with psychosis in a secure forensic psychiatric hospital: a prospective-cohort waiting list controlled study. *BMC Research Notes*, *5*(1), 302. doi: 10.1186/1756-0500-5-302

Novaco, R. W. (1994). Anger as a risk factor for violence among the mentally disordered. In J. Monahan & H. J. Steadman (Eds.), *The John D. and Catherine T. MacArthur Foundation series on mental health and development. Violence and mental disorder: Developments in risk assessment* (p. 21–59). University of Chicago Press.

Novaco, R. W. (2003). *The Novaco Anger Scale and Provocation Inventory: NAS-PI*. Los Angeles, CA: Western Psychological Services.

*O'Reilly, K., Donohoe, G., O'Sullivan, D., Coyle, C., Corvin, A., O'Flynn, P., ... & Kennedy, H. G. (2019). A randomized controlled trial of cognitive remediation for a national cohort of forensic patients with schizophrenia or schizoaffective disorder. *BMC Psychiatry*, *19*(1), 27. doi: 10.1186/s12888-019-2018-6

Overall, J. E., & Gorham, D. R. (1962). The brief psychiatric rating scale. *Psychological reports*, *10*(3), 799-812.

Papalia, N., Spivak, B., Daffern, M., & Ogloff, J. R. (2019). A meta- analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clinical Psychology: Science and Practice*, *26*(2), e12282. doi: 10.1111/cpsp.12282

Pavot, W., & Diener, E. (1993). Review of the life satisfaction scale. *Psychological Assessment*, *5*(2), 162-172.

Platt, J. J. & Spivack, G. (1975). *Manual for the means-end problem solving procedure*. Philadelphia: Department of Mental Health Services, Hahnemann Community Mental Health/Mental Retardation Centre.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

*Rees-Jones, A., Gudjonsson, G., & Young, S. (2012). A multi-site controlled trial of a cognitive skills program for mentally disordered offenders. *BMC Psychiatry*, *12*(1), 44. doi: 10.1186/1471-244X-12-44

Rice, M. E. (1983). Improving the social skills of males in a maximum security psychiatric setting. *Canadian Journal of Behavioural Science*, *15*(1), 1-13. doi: 10.1037/h0080683

Ross, J., Quayle, E., Newman, E., & Tansey, L. (2013). The impact of psychological therapies on violent behaviour in clinical and forensic settings: A systematic review. *Aggression and Violent Behavior*, *18*(6), 761-773. doi: 10.1016/j.avb.2013.09.001

Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press.

Schreurs, P. J. G., Van de Willige, G., Brosschot, J. F., Tellegen, B., & Graus, G. M. H. (1993). *De Utrechtse coping lijst: omgaan met problemen en gebeurtenissen*. [The Utrecht Coping Scale: UCL. Dealing with problems and situations]. Lisse: Swets en Zetlinger.

Schulz, K. F., Altman, D. G., Moher, D., & Consort Group. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Trials*, *11*(1), 32. doi: 10.1186/1745-6215-11-32

Scottish Intercollegiate Guidelines Network (SIGN). (2020a). Algorithm for classifying study design for questions of effectiveness. Available from https://www.sign.ac.uk/assets/study_design.pdf. Last accessed 8 March 2020.

Scottish Intercollegiate Guidelines Network (SIGN). (2020b). Critical appraisal notes and checklists. Available from https://www.sign.ac.uk/checklists-and-notes. Last accessed 8 March 2020.

Simourd, D. J. (1997). The Criminal Sentiments Scale-Modified and Pride in Delinquency scale: Psychometric properties and construct validity of two measures of criminal attitudes. *Criminal Justice and Behavior*, *24*(1), 52-70. doi: 10.1177/0093854897024001004

Sintonen, H. (2001). The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine*, *33*(5), 328-336. doi: 10.3109/07853890109002086

Spielberger, C. D., Sydeman, S. J., Owen, A. E., & Marsh, B. J. (1999). *Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI)*. Lawrence Erlbaum Associates Publishers.

Stermac, L. (1986). Anger control treatment for forensic patients. *Journal of Interpersonal Violence*, *1*(4), 446-457. doi: 10.1177/088626086001004004

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*. doi: 10.1136/bmj.d4002

Sturgeon, M., Tyler, N., & Gannon, T. A. (2018). A systematic review of group work interventions in UK high secure hospitals. *Aggression and Violent Behavior*, *38*, 53-75. doi: 10.1016/j.avb.2017.11.004

Tapp, J., Warren, F., Fife-Schaw, C., Perkins, D., & Moore, E. (2013). What do the experts by experience tell us about 'what works' in high secure forensic inpatient hospital services?. *The Journal of Forensic Psychiatry & Psychology*, *24*(2), 160-178. doi: 10.1080/14789949.2012.760642

Taylor, P.J., Robling, M., Playle, R., Bezeczky, Z., John-Evens, H., Dimitropouluo, P., McNamara, R., Moriarty, Y., Summers, Z., & Bagshaw, R. (in press). A randomised controlled trial of a Group psychological intervention to increase locus of control for alcohol consumption among Alcohol-misusing Short-term (male) Prisoners (GASP). *Addiction*. 10.1111/add.15006.

Thomas, S., McCrone, P., & Fahy, T. (2009). How do psychiatric patients on prison healthcare centres differ from inpatients in secure psychiatric inpatient units? *Psychology, Crime & Law, 15*(8), 720-742. doi: 10.1080/10683160802516265

Tolland, H., McKee, T., Cosgrove, S., Drysdale, E., Gillespie, M., Paterson, L., & Totten, C. (2019). A systematic review of effective therapeutic interventions and management strategies for challenging behaviour in women in forensic mental health settings. *The Journal of Forensic Psychiatry & Psychology*, *30*(4), 570-593. doi: 10.1080/14789949.2019.1627387

Tomlinson, M. F., & Hoaken, P. N. (2017). The potential for a skills-based dialectical behavior therapy program to reduce aggression, anger, and hostility in a Canadian forensic psychiatric sample: a

pilot study. *International Journal of Forensic Mental Health*, *16*(3), 215-226. doi: 10.1080/14999013.2017.1315469

Tong, L. S., & Farrington, D. P. (2006). How effective is the "Reasoning and Rehabilitation" programme in reducing reoffending? A meta-analysis of evaluations in four countries. *Psychology, Crime & Law*, *12*(1), 3-24. doi: 10.1080/10683160512331316253

*Tyler, N., Gannon, T. A., Lockerbie, L., & Ó Ciardha, C. (2018). An evaluation of a specialist firesetting treatment programme for male and female mentally disordered offenders (the FIP- MO). *Clinical Psychology & Psychotherapy*, *25*(3), 388-400. doi: 10.1002/cpp.2172

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistical Software, 36*(3), 1-48. doi: 10.18637/jss.v036.i03

Vojt, G., Slesser, M., Marshall, L., & Thomson, L. (2011). The clinical reality of implementing formal risk assessment and management measures within high secure forensic care. *Medicine, Science and the Law*, *51*(4), 220-227. doi: 10.1258/msl.2011.010149

Völlm, B. A., Edworthy, R., Huband, N., Talbot, E., Majid, S., Holley, J., ... & Duggan, C. (2018). Characteristics and pathways of long-stay patients in high and medium secure settings in England; A secondary publication from a large mixed-methods study. *Frontiers in Psychiatry*, *9*, 140. doi: 10.3389/fpsyt.2018.00140

*Walker, H., Connaughton, J., Wilson, I., & Martin, C. R. (2012). Improving outcomes for psychoses through the use of psycho- education; preliminary findings. *Journal of Psychiatric and Mental Health Nursing*, *19*(10), 881-890. doi: 10.1111/j.1365-2850.2012.01873.x

*Walker, H., Tulloch, L., Ramm, M., Drysdale, E., Steel, A., Martin, C., ... & Connaughton, J. (2013). A randomised controlled trial to explore insight into psychosis; effects of a psycho-education programme on insight in a forensic population. *The Journal of Forensic Psychiatry & Psychology*, *24*(6), 756-771. doi: 10.1080/14789949.2013.853821

*Williams, E., Ferrito, M., & Tapp, J. (2014). Cognitive-behavioural therapy for schizophrenia in a forensic mental health setting. *Journal of Forensic Practice*, *16*(1), 68-77. doi: 10.1108/JFP-12-2012-0028

Woods, P., Reed, V., & Robinson, D. (1999). The Behavioural Status Index: therapeutic assessment of risk, insight, communication and social skills. *Journal of Psychiatric and Mental Health Nursing*, *6*(2), 79-90. doi: 10.1046/j.1365-2850.1999.620079.x

*Yip, V. C., Gudjonsson, G. H., Perkins, D., Doidge, A., Hopkin, G., & Young, S. (2013). A non-randomised controlled trial of the R&R2MHP cognitive skills program in high risk male offenders with severe mental illness. *BMC Psychiatry*, *13*(1), 267. doi: 10.1186/1471-244X-13-267

Yoon, I. A., Slade, K., & Fazel, S. (2017). Outcomes of psychological therapies for prisoners with mental health problems: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology*, *85*(8), 783-802. doi: 10.1037/ccp0000214

*Young, S., Chick K., & Gudjonsson, G. (2010). A preliminary evaluation of reasoning and rehabilitation 2 in mentally disordered offenders (R&R2M) across two secure forensic settings in the United Kingdom. *The Journal of Forensic Psychiatry & Psychology, 21*(3), 336-349. doi: 10.1080/14789940903513203

Young, S., Gudjonsson, G., Ball, S., & Lam, J. (2003). Attention Deficit Hyperactivity Disorder (ADHD) in personality disordered offenders and the association with disruptive behavioural problems. *The Journal of Forensic Psychiatry*, *14*(3), 491-505. doi: 10.1080/14789940310001615461

*Young, S., Hopkin, G., Perkins, D., Farr, C., Doidge, A., & Gudjonsson, G. (2013). A controlled trial of a cognitive skills program for personality-disordered offenders. *Journal of Attention Disorders*, *17*(7), 598-607. doi: 10.1177/1087054711430333

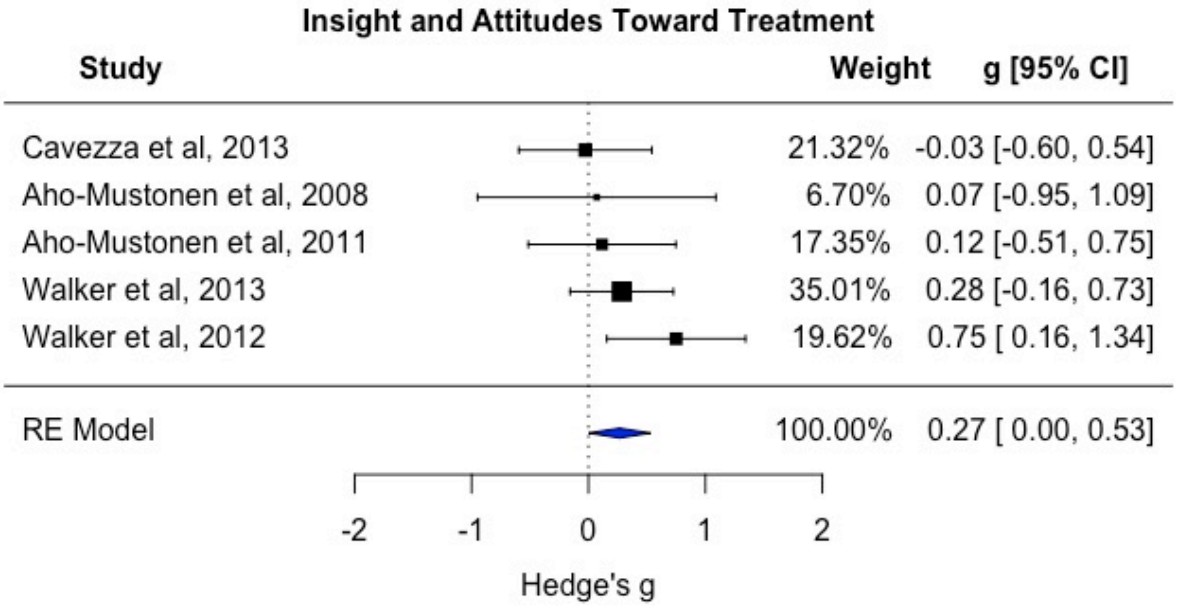
*Zwets, A. J., Hornsveld, R. H., Muris, P., Kanters, T., Langstraat, E., & van Marle, H. J. (2016). Psychomotor therapy as an additive intervention for violent forensic psychiatric inpatients: A pilot study. *International Journal of Forensic Mental Health*, *15*(3), 222-234. doi: 10.1080/14999013.2016.1152613

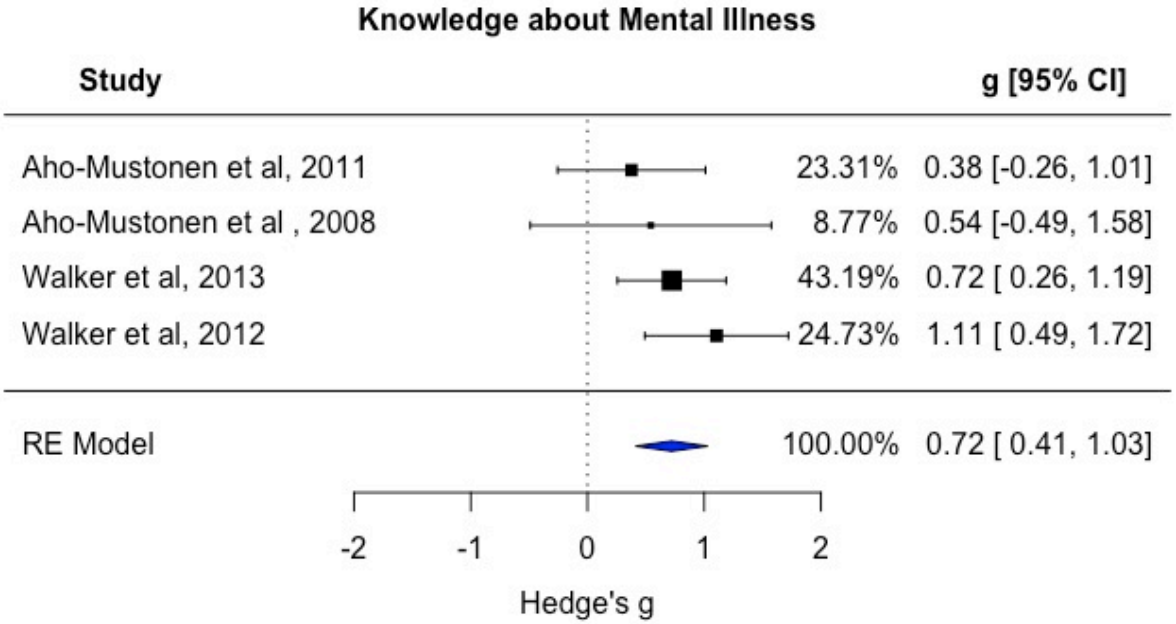
Appendix A

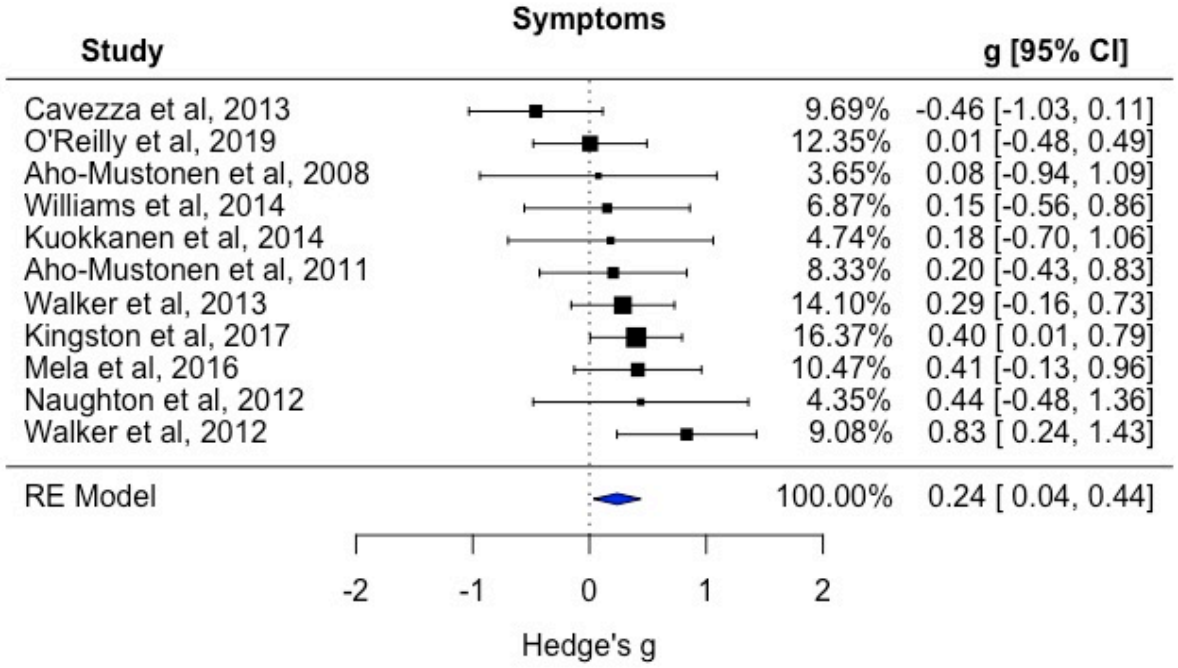
Search keywords for PsycINFO databse

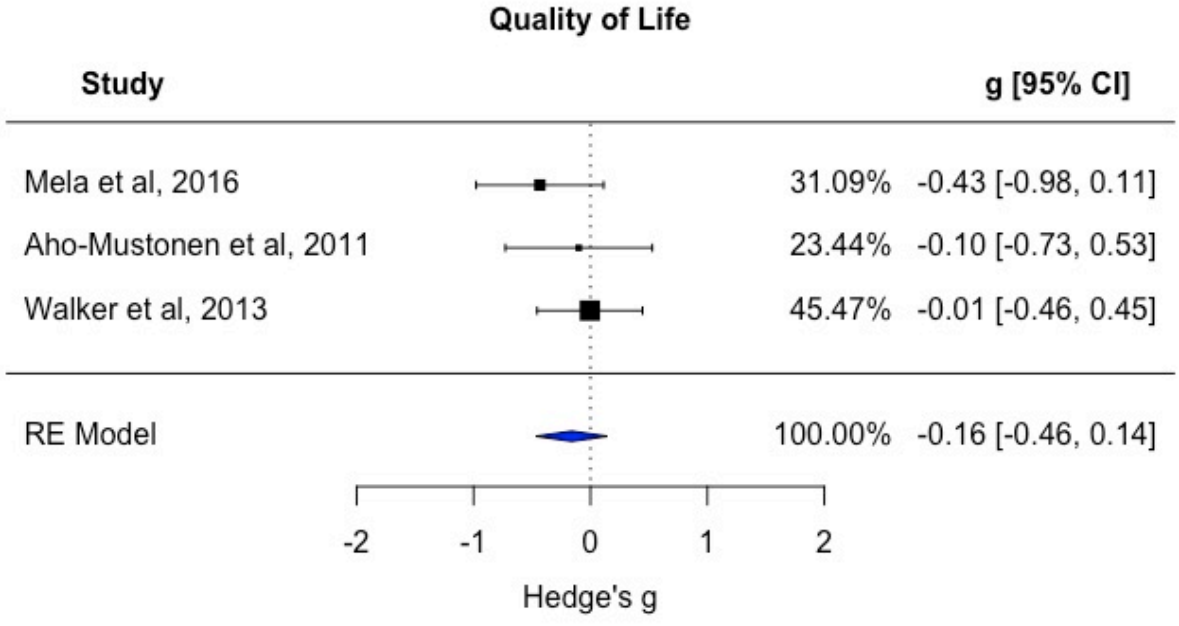
1. (violen* adj2 ?patient*).mp.
2. mental* disord* offend*.mp.
3. (offend* adj2 mental* ill).mp.
4. (inmate* adj2 mental* ill).mp.
5. (offend* adj2 ?patient*).mp.
6. mentally ill offenders/
7. (secur* adj2 ?patient*).mp.
8. secure mental health.mp.
9. (secur* adj2 hospital*).mp.
10. (special adj2 hospital*).mp.
11. (secur* adj2 setting*).mp.
12. (secur* adj2 unit*).mp.
13. (secur* adj2 care).mp.
14. (low adj2 secur*).mp.
15. (medium adj2 secur*).mp.
16. (high adj2 secur*).mp.
17. (forensic adj2 service*).mp.
18. (offend* adj2 ?patient*).mp.
19. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18
20. personalit* disorder*.mp.
21. exp personality disorders/
22. (forensic* or offend*).mp.
23. 20 or 21
24. 22 and 23
25. 19 or 24
26. exp psychotherapy/
27. (cognitiv* adj2 therapy*).mp.
28. (skill* adj2 train*).mp.
29. exp cognitive therapy/ or exp anxiety management/ or exp behavior modification/ or exp cognitive behavior therapy/
30. anger management.mp. or exp Anger Control/
31. intervention*.mp. or exp Response to Intervention/ or exp Group Intervention/ or exp Intervention/
32. program*.mp. or exp Program Evaluation/
33. (therapies or therapy).mp.
34. exp treatment/ or mental health program evaluation/ or posttreatment followup/ or psychoeducation/ or relapse prevention/ or treatment effectiveness evaluation/ or treatment outcomes/
35. 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34
36. 25 and 35
37. limit 36 to (human and ("0400 empirical study" or "0410 experimental replication" or "0430 followup study" or "0450 longitudinal study" or "0451 prospective study" or "0453 retrospective study" or "0800 literature review" or "0830 systematic review" or 1200 meta analysis or 1800 quantitative study or "2000 treatment outcome/clinical trial") and yr="1980-Current")

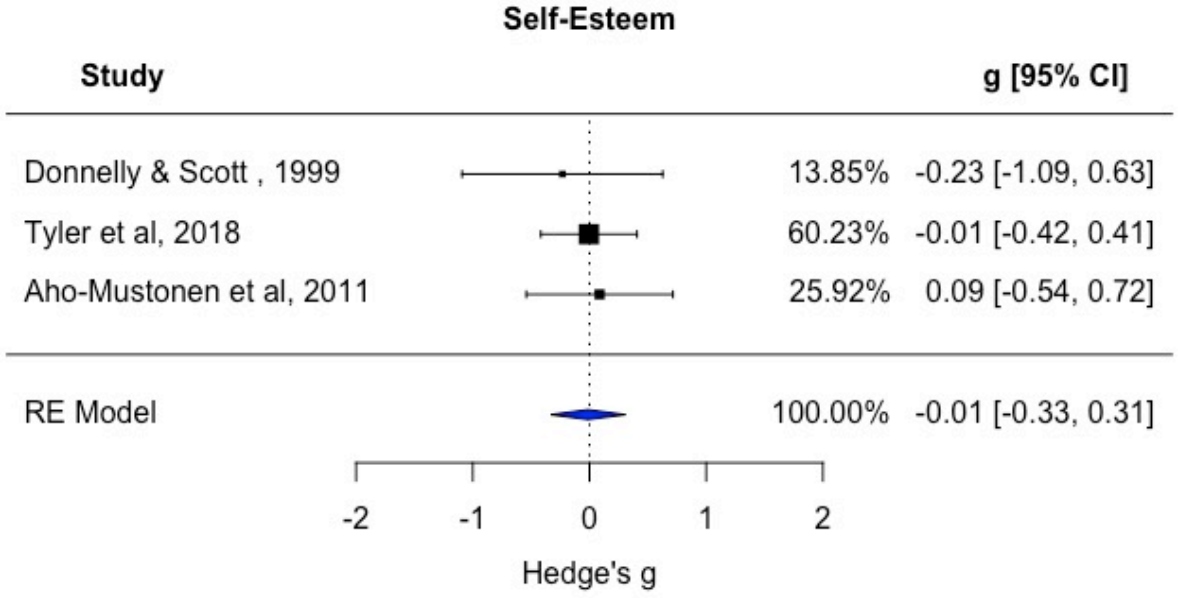
Appendix B
Forest plots for each meta-analysis

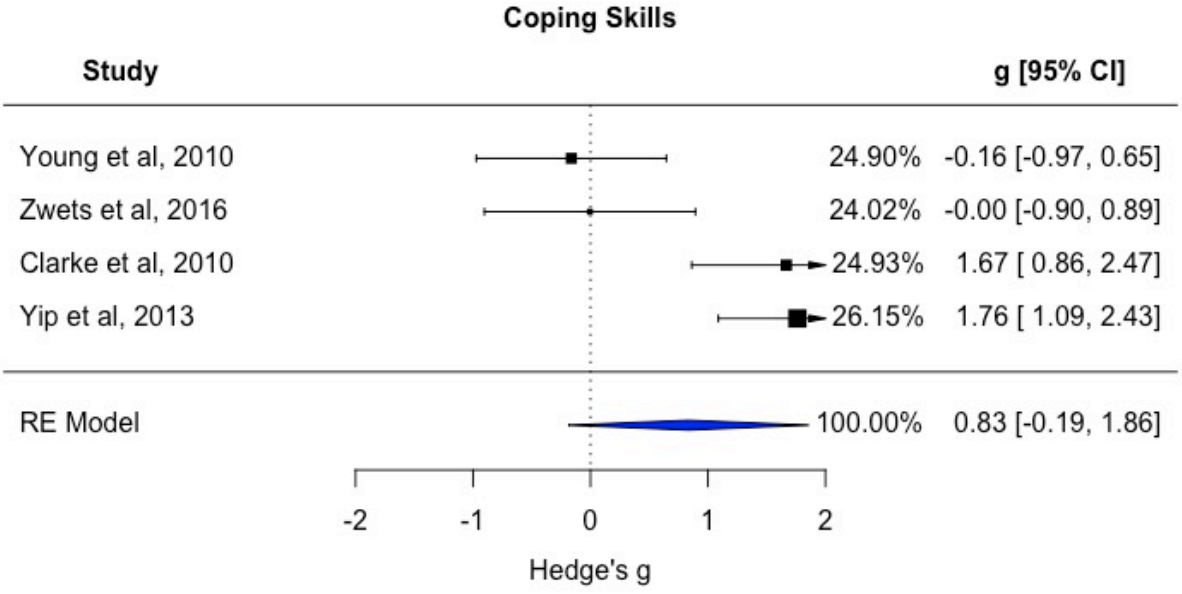


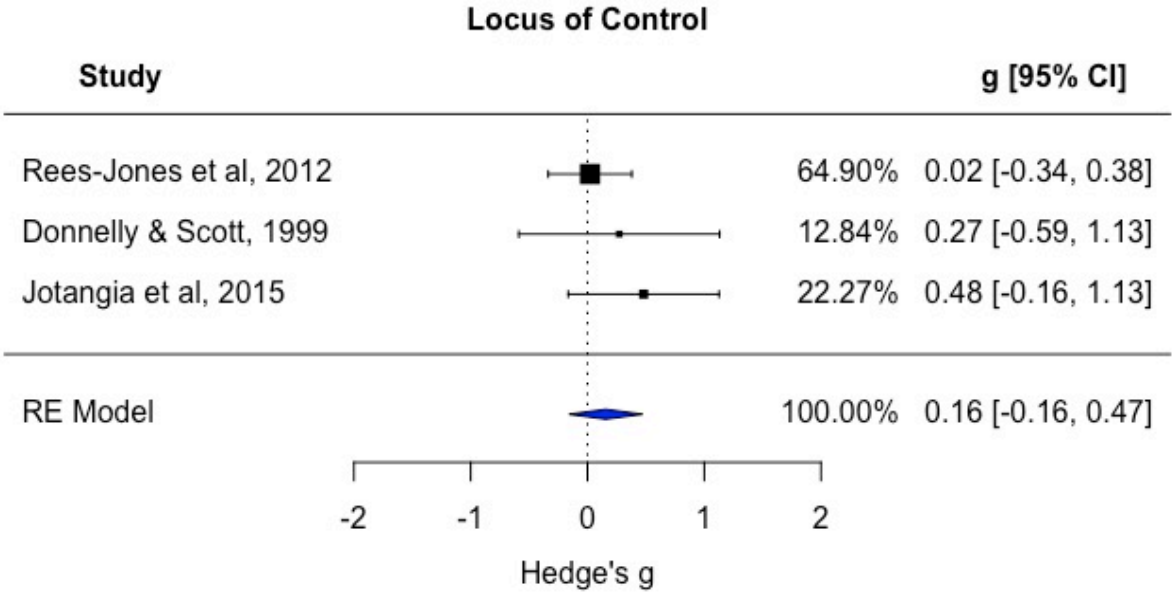


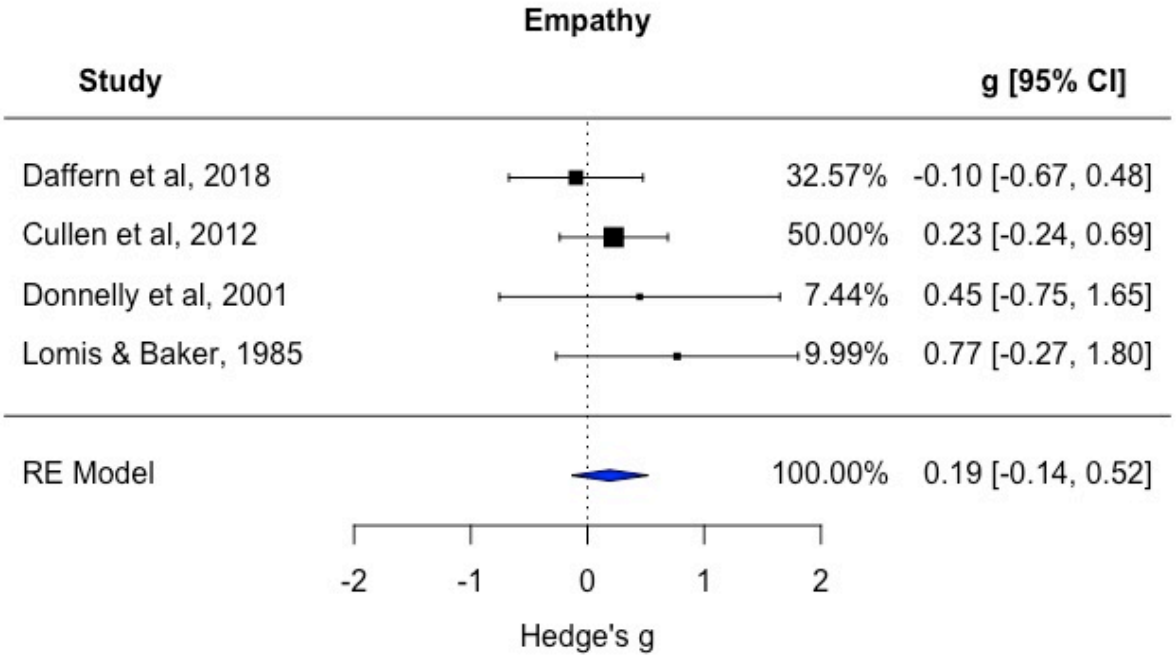

Symptoms
Study
g [95% CI]
Cavezza et al, 2013 9.69% -0.46 [-1.03, 0.11]
O'Reilly et al, 2019 12.35% 0.01 [-0.48, 0.49]
Aho-Mustonen et al, 2008 3.65% 0.08 [-0.94, 1.09]
Williams et al, 2014 6.87% 0.15 [-0.56, 0.86]
Kuokkanen et al, 2014 4.74% 0.18 [-0.70, 1.06]
Aho-Mustonen et al, 2011 8.33% 0.20 [-0.43, 0.83]
Walker et al, 2013 14.10% 0.29 [-0.16, 0.73]
Kingston et al, 2017 16.37% 0.40 [0.01, 0.79]
Mela et al, 2016 10.47% 0.41 [-0.13, 0.96]
Naughton et al, 2012 4.35% 0.44 [-0.48, 1.36]
Walker et al, 2012 9.08% 0.83 [0.24, 1.43]
RE Model 100.00% 0.24 [0.04, 0.44]
-2
-1
0
1
2
Hedge's g

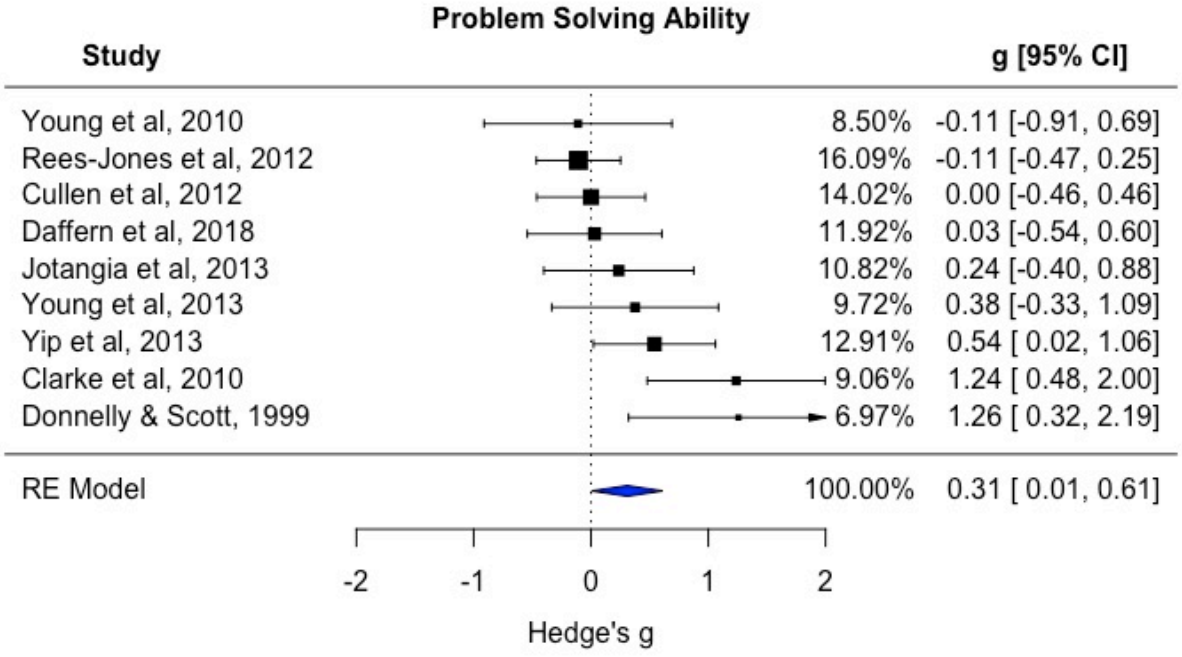

Quality of Life
Study
g [95% CI]
Mela et al, 2016 31.09% -0.43 [-0.98, 0.11]
Aho-Mustonen et al, 2011 23.44% -0.10 [-0.73, 0.53]
Walker et al, 2013 45.47% -0.01 [-0.46, 0.45]
RE Model 100.00% -0.16 [-0.46, 0.14]
-2
-1
0
1
2
Hedge's g

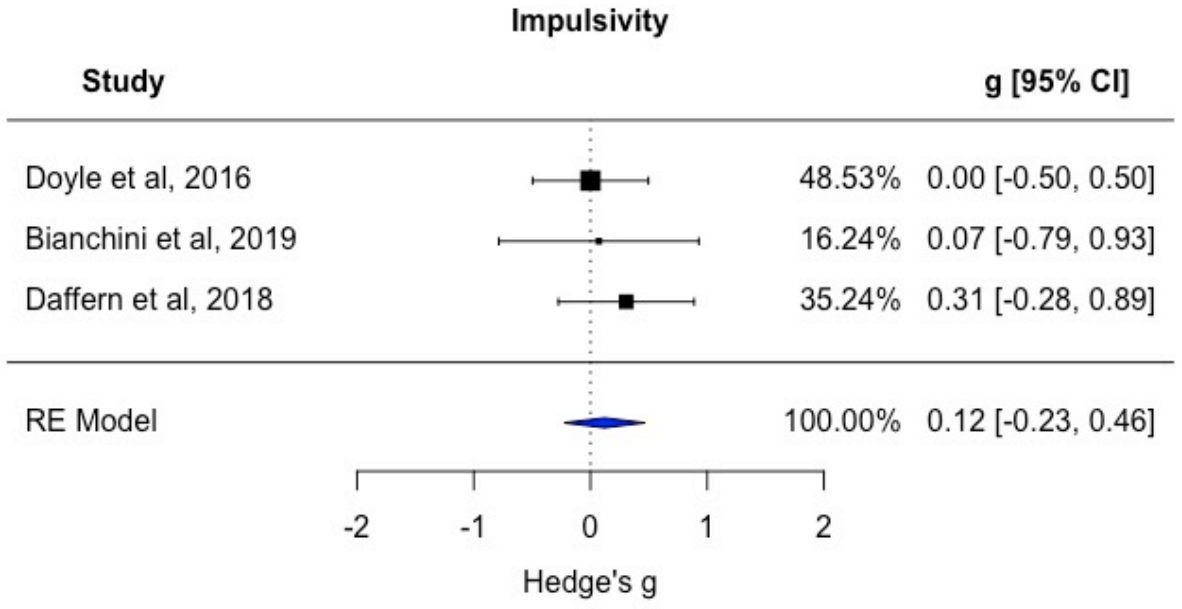

Self-Esteem
Study
g [95% CI]
Donnelly & Scott , 1999
13.85%
-0.23 [-1.09, 0.63]
Tyler et al, 2018
60.23%
-0.01 [-0.42, 0.41]
Aho-Mustonen et al, 2011
25.92%
0.09 [-0.54, 0.72]
RE Model
100.00%
-0.01 [-0.33, 0.31]
-2
-1
0
1
2
Hedge's g

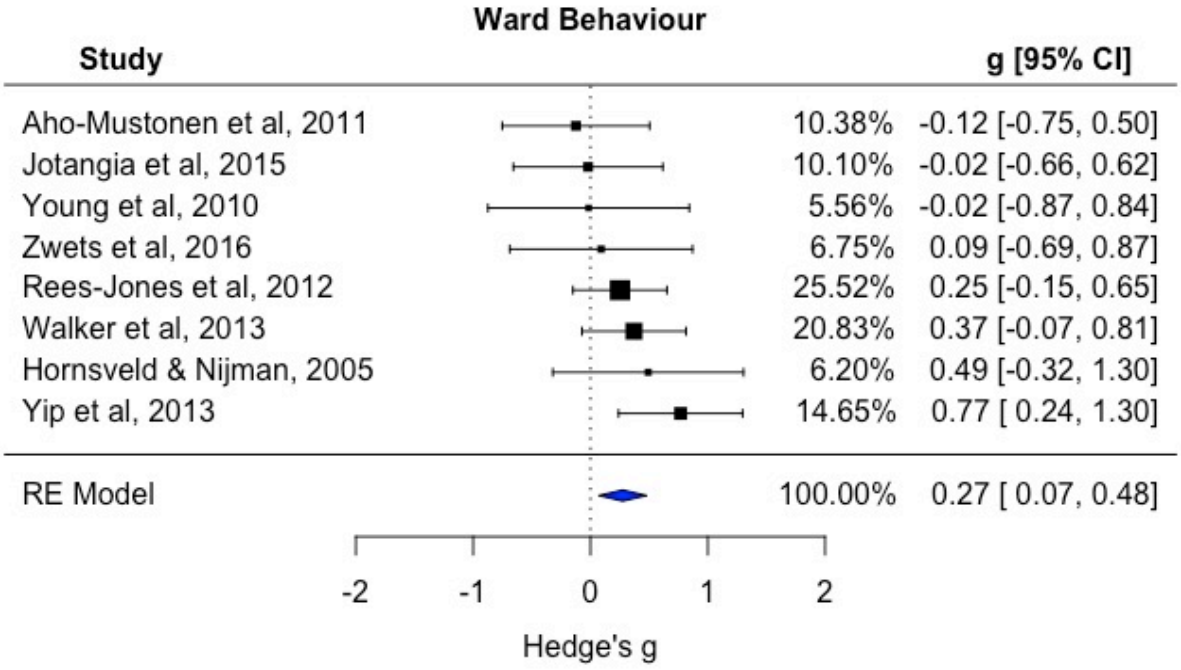

Coping Skills
Study
g [95% CI]
Young et al, 2010
24.90%
-0.16 [-0.97, 0.65]
Zwets et al, 2016
24.02%
-0.00 [-0.90, 0.89]
Clarke et al, 2010
24.93%
1.67 [0.86, 2.47]
Yip et al, 2013
26.15%
1.76 [1.09, 2.43]
RE Model
100.00%
0.83 [-0.19, 1.86]
-2
-1
0
1
2
Hedge's g

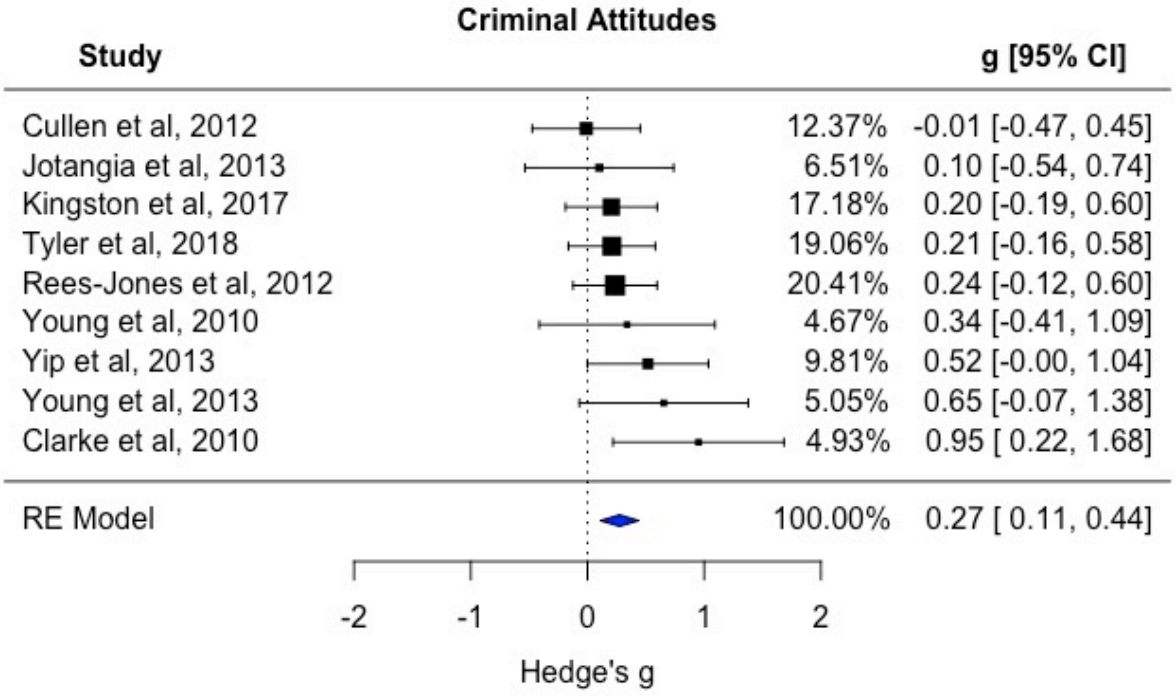

Locus of Control
Study
g [95% CI]
Rees-Jones et al, 2012
64.90%
0.02 [-0.34, 0.38]
Donnelly & Scott, 1999
12.84%
0.27 [-0.59, 1.13]
Jotangia et al, 2015
22.27%
0.48 [-0.16, 1.13]
RE Model
100.00%
0.16 [-0.16, 0.47]
-2
-1
0
1
2
Hedge's g

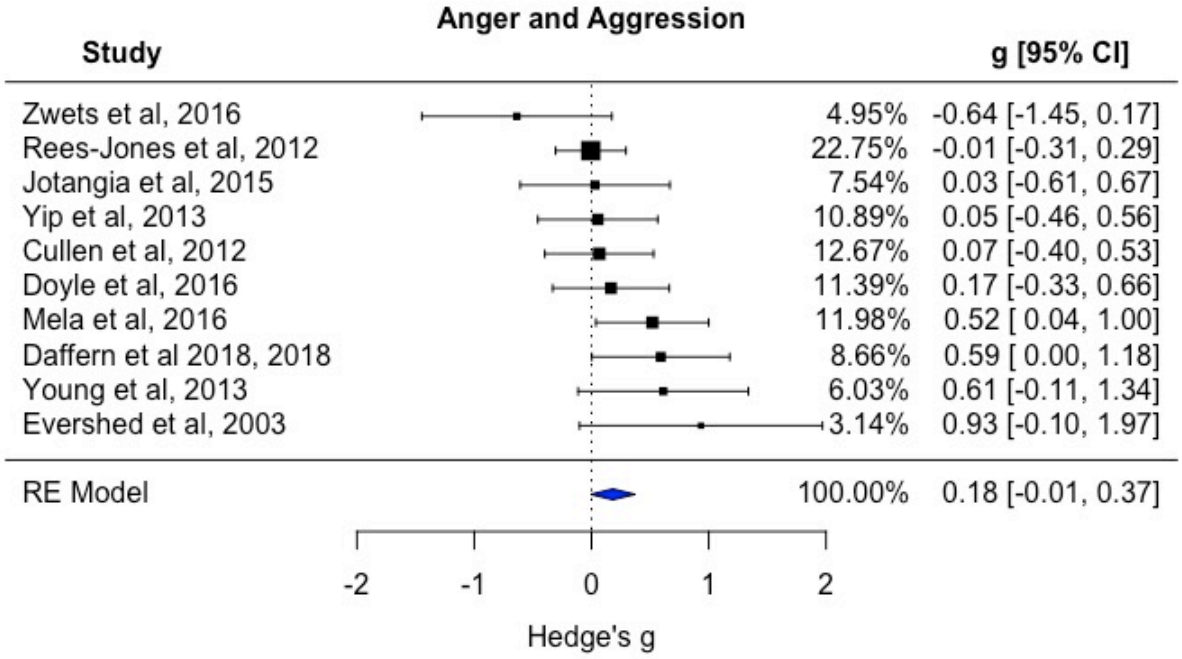

Empathy
Study
g [95% CI]
Daffern et al, 2018
32.57%
-0.10 [-0.67, 0.48]
Cullen et al, 2012
50.00%
0.23 [-0.24, 0.69]
Donnelly et al, 2001
7.44%
0.45 [-0.75, 1.65]
Lomis & Baker, 1985
9.99%
0.77 [-0.27, 1.80]
RE Model
100.00%
0.19 [-0.14, 0.52]
-2
-1
0
1
2
Hedge's g

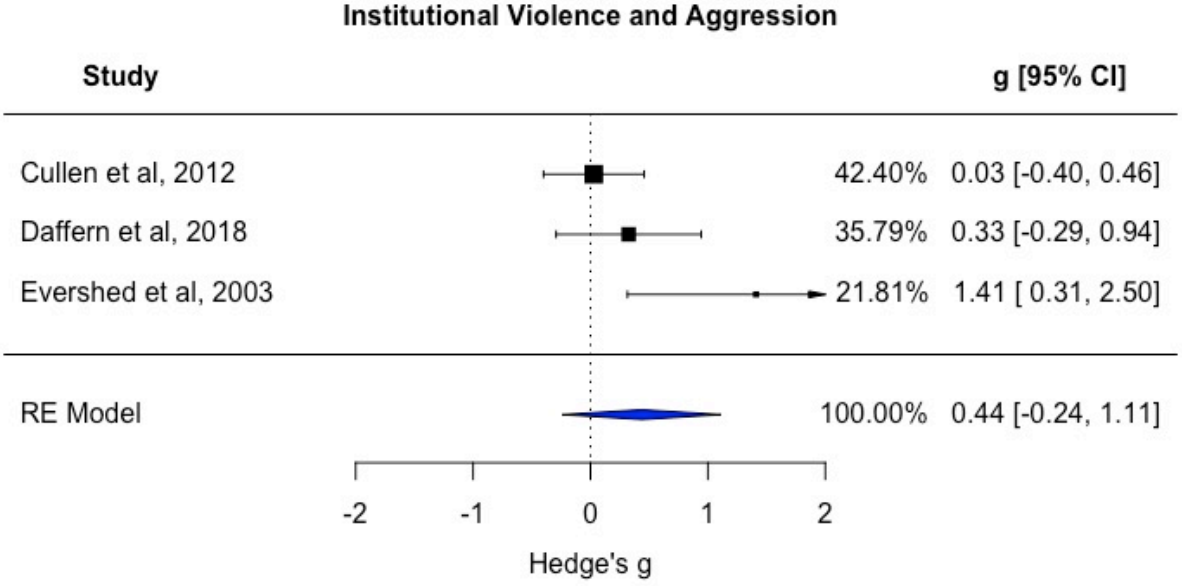

Problem Solving Ability
Study
g [95% CI]
Young et al, 2010 8.50% -0.11 [-0.91, 0.69]
Rees-Jones et al, 2012 16.09% -0.11 [-0.47, 0.25]
Cullen et al, 2012 14.02% 0.00 [-0.46, 0.46]
Daffern et al, 2018 11.92% 0.03 [-0.54, 0.60]
Jotangia et al, 2013 10.82% 0.24 [-0.40, 0.88]
Young et al, 2013 9.72% 0.38 [-0.33, 1.09]
Yip et al, 2013 12.91% 0.54 [0.02, 1.06]
Clarke et al, 2010 9.06% 1.24 [0.48, 2.00]
Donnelly & Scott, 1999 6.97% 1.26 [0.32, 2.19]
RE Model 100.00% 0.31 [0.01, 0.61]
-2
-1
0
1
2
Hedge's g


Impulsivity
Study
g [95% CI]
Doyle et al, 2016 48.53% 0.00 [-0.50, 0.50]
Bianchini et al, 2019 16.24% 0.07 [-0.79, 0.93]
Daffern et al, 2018 35.24% 0.31 [-0.28, 0.89]
RE Model 100.00% 0.12 [-0.23, 0.46]
-2
-1
0
1
2
Hedge's g


Ward Behaviour
Study
g [95% CI]
Aho-Mustonen et al, 2011 10.38% -0.12 [-0.75, 0.50]
Jotangia et al, 2015 10.10% -0.02 [-0.66, 0.62]
Young et al, 2010 5.56% -0.02 [-0.87, 0.84]
Zwets et al, 2016 6.75% 0.09 [-0.69, 0.87]
Rees-Jones et al, 2012 25.52% 0.25 [-0.15, 0.65]
Walker et al, 2013 20.83% 0.37 [-0.07, 0.81]
Hornsveld & Nijman, 2005 6.20% 0.49 [-0.32, 1.30]
Yip et al, 2013 14.65% 0.77 [0.24, 1.30]
RE Model 100.00% 0.27 [0.07, 0.48]
-2
-1
0
1
2
Hedge's g


Criminal Attitudes
Study
g [95% CI]
Cullen et al, 2012 12.37% -0.01 [-0.47, 0.45]
Jotangia et al, 2013 6.51% 0.10 [-0.54, 0.74]
Kingston et al, 2017 17.18% 0.20 [-0.19, 0.60]
Tyler et al, 2018 19.06% 0.21 [-0.16, 0.58]
Rees-Jones et al, 2012 20.41% 0.24 [-0.12, 0.60]
Young et al, 2010 4.67% 0.34 [-0.41, 1.09]
Yip et al, 2013 9.81% 0.52 [-0.00, 1.04]
Young et al, 2013 5.05% 0.65 [-0.07, 1.38]
Clarke et al, 2010 4.93% 0.95 [0.22, 1.68]
RE Model 100.00% 0.27 [0.11, 0.44]
-2
-1
0
1
2
Hedge's g


Anger and Aggression
Study
g [95% CI]
Zwets et al, 2016 4.95% -0.64 [-1.45, 0.17]
Rees-Jones et al, 2012 22.75% -0.01 [-0.31, 0.29]
Jotangia et al, 2015 7.54% 0.03 [-0.61, 0.67]
Yip et al, 2013 10.89% 0.05 [-0.46, 0.56]
Cullen et al, 2012 12.67% 0.07 [-0.40, 0.53]
Doyle et al, 2016 11.39% 0.17 [-0.33, 0.66]
Mela et al, 2016 11.98% 0.52 [0.04, 1.00]
Daffern et al 2018, 2018 8.66% 0.59 [0.00, 1.18]
Young et al, 2013 6.03% 0.61 [-0.11, 1.34]
Evershed et al, 2003 3.14% 0.93 [-0.10, 1.97]
RE Model 100.00% 0.18 [-0.01, 0.37]
-2
-1
0
1
2
Hedge's g


Institutional Violence and Aggression
Study
g [95% CI]
Cullen et al, 2012 42.40% 0.03 [-0.40, 0.46]
Daffern et al, 2018 35.79% 0.33 [-0.29, 0.94]
Evershed et al, 2003 21.81% 1.41 [0.31, 2.50]
RE Model 100.00% 0.44 [-0.24, 1.11]
-2
-1
0
1
2
Hedge's g

Appendix C

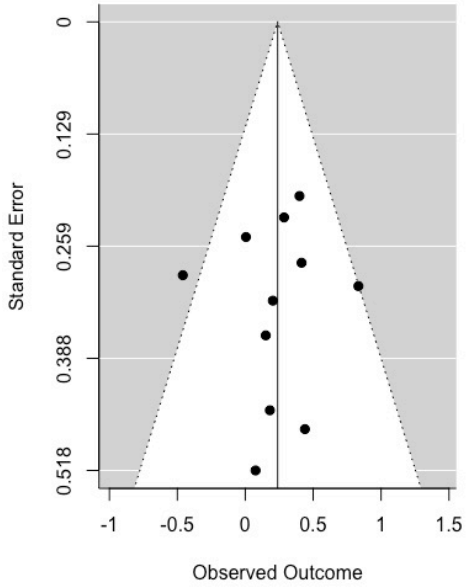


Figure 1. Funnel plot analysis of studies reporting on symptoms (k = 11)

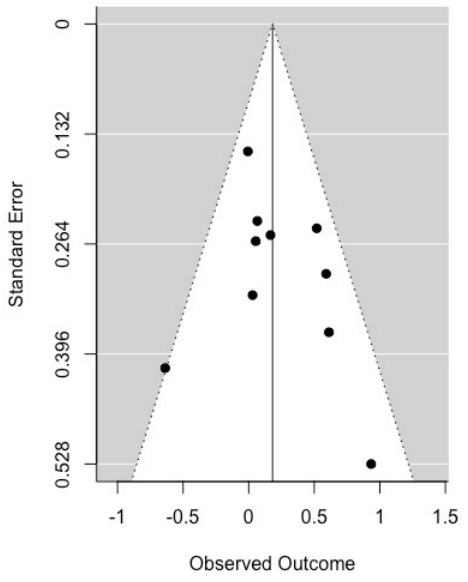


Figure 2. Funnel plot analysis of studies reporting on anger and aggression (k = 10)