Trauma and Treatment Interference in Forensic Populations

by

Annie Sillence

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF CLINICAL PSYCHOLOGY

Centre for Applied Psychology
School of Psychology
College of Life and Environmental Sciences
University of Birmingham

April 2024

Thesis Overview

This thesis comprises four chapters. The first chapter is a meta-analysis of the prevalence of post traumatic stress disorder (PTSD) in female prison populations. Twenty articles were included, reporting a total of 7266 participants. The average PTSD prevalence was 24%. Sensitivity analyses found lower rates of prevalence when diagnostic interviews were conducted by psychiatrists and clinical psychologists, compared to other professionals. The risk of bias found in some of the studies highlights a need for future research to use more rigorous methodology to minimise risk of bias. Recommendations are made for increasing the diagnosis and treatment of PTSD in prisons to improve outcomes for those living with PTSD.

The second chapter is an exploratory analysis of predictors of successful and non-successful outcomes at HMP Grendon. A series of binary logistic regressions were conducted, and a clinical regression decision tree was calculated. The clinical regression tree was able to predict non-successful outcomes with an overall accuracy of 97.6%, but successful outcomes with a 7.9% accuracy. Men who self-report high Antisocial Features, Treatment Rejection and Dominance, and low Social Self-Esteem, have a 98.1% chance of a non-successful outcome. This predictive model holds significant clinical implications, enabling the identification of treatment-interfering factors. This information could be used to adjust treatment accordingly and reduce the number of people that leave Grendon with a non-successful outcome. Further research is required to understand common psychometric profiles that accurately predict successful outcomes.

The third and fourth chapters consist of press releases for the first and second chapters.

Dedications

I would like to dedicate this thesis to my wonderful family and friends. Without your support and encouragement this would never have been possible. Thank you for always believing in me.

Acknowledgements

I would like to thank my brilliant supervisors, Dr Chris Jones, Dr Sai Tung Ho and Dr Stephanie Wilson for their endless advice, guidance, and words of wisdom. I am eternally grateful for all their support.

Thank you to the staff at HMP Grendon, for their encouragement and expertise, and showing me so much kindness. I am so impressed by what is achieved at HMP Grendon and wish more services like this existed.

A final thank you to all the residents at HMP Grendon for talking to me and making me feel so welcome. It was a privilege to work with them, and I wish them all the best on their continued journey.

Table of Contents

List of Figures

List of Tables

Chapter 1: The prevalence of post traumatic stress disorder in female prison populations

Abstract

Introduction

It is well established that women in the criminal justice system have been exposed to a high level of danger throughout their lives (Byrne & Howells, 2000; Fleming et al., 2001). Previous research has found a point prevalence of post traumatic stress disorder (PTSD) in female prison populations of 21.1% (Baranyi et al., 2018). However, there have been significant changes to PTSD diagnostic criteria in recent years. This meta-analysis will aim to update the prevalence rate of PTSD in female prison populations.

Methods

A systematic search of the literature identified a total of 1296 articles. 20 articles were included, reporting 7266 participants. All studies were assessed for risk of bias. Event rates were pooled using a random-effects meta-analysis, and sensitivity analyses were conducted.

Results

Pooled point prevalence of PTSD was 22% and pooled period prevalence was 29%. Sensitivity analyses revealed significantly lower rates of prevalence in studies where

diagnostic interviews were conducted by psychiatrists and clinical psychologists, compared to other professionals or trained interviewers.

Discussion

Women experience substantially higher rates of PTSD, when compared to prevalence rates in the literature for male offenders, and community samples (Baranyi et al., 2018; Koenen et al., 2017). Increasing diagnosis and treatment of PTSD in prisons could improve quality of life of individuals, and reduce offending rates (Facer-Irwin et al., 2019; Sadeh & McNiel, 2015). A wide scale implementation of trauma-informed approaches must be conducted and evaluated to improve outcomes for women in prison (Levenson & Willis, 2019; Miller & Najavits, 2012).

Introduction

Female offenders constitute 6.9% of the global prison population (Fair & Walmsley, 2022), but this number is increasing at an alarming rate. The number of women and girls in prison has increased by nearly 60% since the year 2000, compared with a 22% increase in men (Fair & Walmsley, 2022). The majority of women in prison face complex social problems that have contributed to their incarceration, such as addiction, sexual and physical abuse, low educational attainment, unemployment, mental and physical health problems (Medlicott, 2012). It is estimated that about 75% of women in prisons are mothers, and due to their frequent role as primary carers, an increase in female incarceration is placing greater strain on their families, and society, due to increased need for foster care placements and adverse outcomes for children (Katz, 2017; Thompson & Harm, 2000). In 2007, a report

commissioned by the United Kingdom (UK) Government concluded that vulnerable women are frequently being incarcerated for minor, non-violent offences in which incarceration is both inappropriate and disproportionate (Corston, 2007). Worldwide crime statistics reveal that women are most likely to be incarcerated for non-violent crimes such as drugs or property crimes (such as theft, fraud or forgery). In comparison, men are more likely to be incarcerated for violent crimes such as rape or murder (Heiskanen & Lietonen, 2016). The Corston report highlights society's lack of sympathy for these women who often end up in prison for committing minor offences, after a lifetime of abuse and mental illness (Corston, 2007). The report calls for a change to the way women are treated in the criminal justice system. It recommends that alternative approaches such as women's centres and community orders should be utilised more frequently, particularly for women who have been exposed to high levels of traumatic experiences, and have committed non-violent crimes and pose little risk to the public (Corston, 2007).

It is well established in research literature that individuals in prison have experienced more danger than the general population, and subsequently rates of traumatic exposure are higher (Bowler et al., 2018; Dalsklev et al., 2021; Vitopoulos et al., 2019). Research exploring Adverse Childhood Experiences (ACE), such as physical, emotional or sexual abuse, witnessing domestic abuse, or parental incarceration and parental substance misuse or mental health difficulties, reveal that women who are incarcerated experience significantly more ACEs that men who are incarcerated, or the general population (Messina et al, 2007; Grella et al., 2013). A study of 2000 prisoners in Minnesota found that 57% of women who are incarcerated reported 4 or more ACEs (Clark & Duwe, 2024), compared to the general population of whom 61% report either exposure to 0 or 1 ACEs, and only 15% report 4 or

more ACEs (Merrick et al., 2018). Women in prison report a higher incidence of childhood abuse and neglect than men in prison (Armytage et al., 2000; Fleming et al., 2001), and are more likely to experience mental health problems (James & Glaze, 2006). Furthermore, the rates of sexual abuse are so high in incarcerated women that some researchers have suggested it may be a critical pathway to prison for women, through the development of mental illness and substance abuse (Karlsson & Zielinski, 2020). Once incarcerated, offenders are also more likely to be exposed to dangerous situations in prison, including physical, sexual, or psychological abuse, solitary confinement, or denial of food (Dierkhising, 2014; Maschi et al., 2015).

Due to the increased levels of trauma exposure throughout their lifetime, it is unsurprising that people in prison experience higher rates of post traumatic stress disorder (PTSD) than the general population (Fovet et al., 2023). Based on a large-scale cross-national study, the point prevalence of PTSD in community samples has been found to vary from 0.3 to 1.9% depending on the country (Koenen et al., 2017). However, it is substantially higher in prisons. Baranyi et al. (2018) conducted a meta-analysis and found that point prevalence of PTSD was 6.2% for male offenders, and 21.1% for female offenders. Research indicates that interpersonal trauma, such as sexual abuse is more likely to lead to PTSD than non-interpersonal trauma (Byrne & Howells, 2000; Forbes et al., 2013; O'Donnell et al., 2017). Interpersonal trauma is experienced more frequently by females than males, and this may explain, in part, why research indicates that PTSD is more prevalent in female offending populations than male offending populations (Baranyi et al., 2018; Forbes et al., 2013).

Research indicates that PTSD is often undiagnosed and untreated in prison populations (Jakobowitz et al., 2017; Tyler et al., 2019). The presence of PTSD may have a detrimental impact on the individual's ability to benefit from rehabilitative programmes offered in prisons (Allely & Allely, 2020). Research has also identified an association between PTSD and higher rates of re-offending (Karatizias et al., 2018; Kubiak, 2004). People with PTSD in prisons are also at a higher risk of co-morbid depressive, anxiety, or substance use disorder diagnoses (Facer-Irwin et al., 2019). Additionally, they are more likely to experience suicidality, self-harm, and aggressive behaviour (Facer-Irwin et al., 2019). Furthermore, PTSD symptoms have been found to be significantly associated with violence in prisons (McCallum, 2018). Research indicates that individuals in the prison system who are experiencing PTSD can benefit from trauma informed interventions, and the widespread implementation of trauma informed care (King, 2017; Levenson & Willis, 2018) A review of 9 studies examining trauma informed interventions found significant reductions in PTSD symptoms, anger and aggression, and improvements in mental health (King, 2017). Furthermore, a longitudinal study conducted by Lehrer (2001) found strong evidence of reduced recidivism rates for incarcerated women who took part in trauma-informed programmes compared to those who did not. However, the current service provision for trauma-focussed interventions in the prison system is poor, and needs to be urgently addressed (Campbell et al., 2016; Pettus-Davis et al., 2019). There are still significant gaps in our understanding of PTSD in prisons, and further research is required to understand this population, and increase the rates of diagnosis, treatment, and trauma informed care (Allely & Allely, 2020; Baranyi et al., 2018; Prost et al., 2022).

Diagnostic criteria

There are currently significant differences in the criteria used to diagnose PTSD in the Diagnostic and Statistical Manual Fifth edition (DSM-5) (American Psychiatric Association, 2022) and the eleventh edition of the International Classification of Diseases (ICD-11) (World Health Organisation, 2021). The DSM-5 categorises PTSD as a broad diagnosis, and included significant revisions to the criteria for PTSD, compared with the DSM-IV (American Psychiatric Association, 2000, 2022). Changes to the diagnostic criteria include alterations to the definition of trauma, increase and rearrangement of the symptoms, and the creation of additional symptoms and criteria (Pai et al., 2017). The DSM-5 characterises PTSD as a disorder in which an individual experiences a range of symptoms from 4 distinct diagnostic clusters: re-experiencing (such as flashbacks or nightmares), avoidance (such as avoiding thoughts, memories, feelings or external reminders of the distressing event), negative cognitions and mood (such as diminished interest in activities, blaming self for what happened), and arousal symptoms (such as hypervigilance, sleep problems, aggressive behaviour) (American Psychiatric Association, 2022). New symptoms added to the PTSD criteria in the DSM-5 are persistent negative emotional state, persistent distorted cognitions about the cause or consequences of the trauma leading to blame of self or others, and reckless or self-destructive behaviour (American Psychiatric Association, 2022).

In contrast to the DSM-5, the ICD-11 has also significantly revised its criteria for PTSD, and now further differs from the DSM. As well as omission of some of the existing symptoms such as sleep disturbances and concentration difficulties, the ICD-11 now conceptualises PTSD as two ‘sibling’ disorders, PTSD and Complex PTSD (C-PTSD) (WHO, 2021). C-PTSD was included as an additional classification of PTSD after decades of clinical

observations that repeated chronic exposure to traumatic experiences (such as childhood sexual abuse, domestic abuse, torture or war) can lead to complex trauma reactions and changes in self-organisation (Herman, 1992; Robles et al., 2014). C-PTSD diagnostic criteria includes the existing primary elements of PTSD, alongside three additional clusters of symptoms related to disturbances in self-organisation: emotional dysregulation, negative self-beliefs, and interpersonal difficulties (WHO, 2021). This additional classification of C-PTSD is highly debated, and is not present within the DSM-5. The DSM-5 took a different path, and included some of the 'complex' symptoms that are found within C-PTSD, within the main PTSD diagnostic criteria (American Psychiatric Association, 2022; Friedman, 2013).

Consequently, there are currently significant differences in diagnostic criteria used for PTSD (Hyland et al., 2018). Research indicates that congruence between the prevalence rates of PTSD of the ICD-11 and DSM-5 are exceptionally low, with the ICD-11 criteria identifying significantly less instances of PTSD than the DSM-5 (Hafstad et al., 2017). Due to the expansion of new symptom presentations, there are now over half a million combinations of symptom profiles possible under the same diagnosis of PTSD within the DSM-5 alone, let alone with the additional variance added by the ICD-11 criteria (Galatzer-Levy and Bryant, 2013). This staggeringly high level of variance in the PTSD diagnostic profile is likely to lead to heterogeneity in the research and may partly explain why findings concerning PTSD are often inconsistent (Galatzer-Levy & Bryant, 2013). Inconsistencies have been found not only amongst prevalence rates (O'Connor et al., 2007), but also amongst effect sizes for treatment outcomes (Bisson et al., 2007) and structural brain abnormalities associated with PTSD (Karl et al., 2006). There is evidence that the significant differences in diagnostic criteria in the DSM-5 and ICD-11 will have negative implications for research, and further increase the heterogeneity of this population (Facer-Irwin et al., 2022; Firmin et al., 2015). Due to the

recent changes to diagnostic criteria, this review will aim to provide an updated calculation of the prevalence of PTSD in female offending populations. Sensitivity analysis will be conducted to assess whether there are differences in prevalence rates due to the differing diagnostic criteria used.

Measurement tools

To add further heterogeneity to this population, there is a wide variety of measurement tools used in clinical practice and research to identify the presence of PTSD. Research indicates that there is only a 45% similarity score between PTSD measurement tools, which indicates that this may contribute to an increase in variance in this population (Newson et al., 2020). Two common categories of tools are often used to identify the presence of PTSD: self-report screening questionnaires and structured clinical interviews. Research indicates that both forms of assessing PTSD are frequently used in prisons (Dulisse et al., 2023). There are a wide variety of self-report screening tools used to establish a possible diagnosis of PTSD, however these methods are only indicative, and a diagnosis of PTSD cannot be made without a clinical interview (National Institute for Health and Care Excellence, 2018). Self-report measures are subject to response bias, misinterpretation, and contextual factors that can introduce significant measurement error (McDonald & Calhoun, 2010). To assist clinicians in assessing and diagnosing PTSD, structured clinical interviews are used both in research and clinical settings to provide a comprehensive evaluation of symptoms (Summerfeldt & Antony, 2002). There is evidence to suggest that clinically administered PTSD scales perform more accurately, as self-report measures tend to overestimate a PTSD diagnosis (Forbes et al., 2001; Griffin et al., 2004). A variety of structured interviews can be used to determine a

diagnosis of PTSD, including: The Clinically Administered PTSD Scale for DSM-5 (CAPS-5) (Weathers et al., 2013) which has widely been hailed the 'gold standard' for PTSD diagnosis (Hunt et al., 2018), the Structured Clinical Interview for DSM-5 (SCID) (First et al., 2015), the Mini Neuropsychiatric Interview (MINI) (Sheehan et al., 1998) and the Composite International Diagnostic Interview (CIDI) (WHO, 1990). To establish an accurate estimation of the prevalence of PTSD that meets diagnostic criteria within female prison populations, this review will be limited to studies that used a structured clinical interview rather than a self-report screening measure.

International landscape

Prison populations around the world vary significantly in size and composition, influenced by diverse factors such as legal systems, crime rates, socio-economic conditions, and political structures (Coyle & Fair, 2013). The United States, Thailand, and El Salvador have the highest rates of incarcerated females, while African countries and some European countries such as Iceland and Norway have much lower rates (Fair & Walmsley, 2022). These disparities are likely to be shaped by differences in legal frameworks, policing practices, gender norms, sentencing laws, and cultural attitudes towards crime and punishment (Cavadino & Dignan, 2006).

The pathways into prison for women with PTSD can differ markedly based on regional and national contexts. As previously discussed, women with PTSD often enter the criminal justice system through behaviours associated with trauma, such as substance abuse, involvement in abusive relationships, or crimes of survival (e.g., theft, sex work) (Covington, 2018). In countries with high incarceration rates, such as the United States, these pathways may be

exacerbated by systemic issues such as lack of mental health services and socio-economic support (Cloud et al., 2014). In contrast, countries with more robust social support systems and rehabilitative justice models may offer alternative pathways to reduce the likelihood of an individual with PTSD being incarcerated. For instance, in Nordic countries the criminal justice system frequently implements diversion programs and community-based sentences that prioritise treatment over incarceration (MacDonald, 2013). These approaches can mitigate the risk of imprisonment for women with PTSD by addressing the underlying causes of criminal behaviour.

It should be noted that the understanding, diagnosis, and treatment of PTSD in both forensic and community settings will differ substantially across different countries and cultures included in this meta-analysis. There are various cultural, racial and ethical considerations that must be incorporated into the assessment of PTSD. Research indicates that some cultures may experience disproportionate exposure to trauma, and an increased risk of PTSD due to factors such as racism, intergenerational trauma, war, environmental disasters, and sociopolitical events (Kerig, et al, 2020). Although there is evidence of cross-cultural validity of PTSD internationally, prevalence rates differ substantially, and a previous international meta-analysis in female international prison populations found a range from 10.6% in China, to 37.6% in Australia (Baranyi et al., 2018). It is important to note that as well as varying exposure to trauma, cultural understanding and explanations of trauma symptoms are also substantially different internationally and will impact subsequent prevalence rates found in this current meta-analysis.

Aims of the current review

This meta-analysis aims to update the literature and establish the current prevalence of PTSD in female prison populations. Due to the changing diagnostic criteria used to diagnose PTSD, and an increase in research in recent years, it is imperative that prevalence rates are reviewed and updated accordingly (American Psychiatric Association, 2022; WHO, 2021). This meta-analysis will focus on reviewing the literature to establish point prevalence of PTSD, and period prevalence of 6 and 12 months. Lifetime prevalence of PTSD will not be reviewed, as we want to establish the rate of PTSD currently within the forensic population, rather than a risk of lifetime prevalence of the disorder. Due to limited research into C-PTSD in female prison populations, this meta-analysis will focus only on PTSD. It is recommended that further research should be conducted into the prevalence of C-PTSD in female prison populations before a review in undertaken. This meta-analysis will include female prison populations that use a structured clinical interview to diagnose PTSD, using either DSM or ICD criteria. Sensitivity analysis will be conducted to examine sources of heterogeneity within the literature.

Methods

This meta-analysis was produced using The University of Birmingham Applied Psychology evidence synthesis protocols, which includes standardised methods and description for the selection of the meta-analytic model, the omnibus test, identification of influential studies,

risk of bias analysis and identification of other covariates and publication bias and small study effects.

Identifying primary studies

Search of Electronic Databases

A systematic search of the literature was conducted in June 2023. The search consisted of the following online databases: Medline, PubMed, Web of Science, PsycInfo and ProQuest (Psychology and Nursing and Allied Healthcare databases). Table 1.1 contains details of the search terms used to identify these studies.

Table 1.1

Search terms used in the systematic search of the literature

Construct	**Free Text Search Terms**	**Method of Search**	**Limits**
Post Traumatic Stress Disorder	"Post traumatic stress disorder" "PTSD" "Posttraumatic stress disorder" "Complex post traumatic stress disorder*" "Complex PTSD" "C-PTSD" "CPTSD"	Free search terms All search terms combined with *OR*	Peer reviewed articles English language
AND			
Prison settings	"Correctional" "Custod*" "Detain*" "Gaol"	Free search terms All search terms combined with *OR*	Peer reviewed articles English language

	"Forensic" "Prison" "Jail" "Inmate" "Offen*" "Penal" "Incarc*" "Imprison*" "Remand" "Sentenced"		
AND			
Prevalence	"Prevalan*" "Epidemiolog*" "Population*" "Incidence"	Free search terms All search terms combined with *OR*	Peer reviewed articles English language
AND			
Sex	"Wom?n" "Female"	Free search terms All search terms combined with *OR*	Peer reviewed articles English language

Inclusion Criteria

Inclusion criteria included studies reporting data from female participants, who were incarcerated in prison. PTSD diagnosis was established using the DSM or the ICD and was collected using validated instruments as part of a clinical or research interview. The full exclusion and inclusion criteria are described below in Table 1.2.

Table 1.2

Inclusion and exclusion criteria

Inclusion criteria		Justification
1.	Data was collected from the general prison population	Prevalence across the general prison population was desired to ascertain the PTSD prevalence in this demographic. Therefore, other populations such as community or military samples were not included
2.	Data was collected from female prisoners	Data from female prisoners was collected to establish prevalence among female prison population, due to research indicating that it is more prevalent than in male prison populations. If both male and female data was collected in the study, then female data was extracted if it was reported separately.
3.	The PTSD prevalence in the sample was reported	This data was required to conduct the meta-analysis on prevalence
4.	The PTSD diagnosis met the criteria as defined by the Diagnostic and Statistical Manual of Mental Disorders, or the International Classification of Diseases	Diagnosed PTSD as defined by the DSM or ICD was sought, to determine prevalence of PTSD, rather than prevalence of possible PTSD which may be indicated by using screening measures
5.	Data concerning a PTSD diagnosis was collected using validated instruments as part of a clinical or research interview	Diagnosed PTSD as identified by clinically administered interviews was sought, to determine prevalence of PTSD, rather than prevalence of possible PTSD which may be indicated by using screening measures

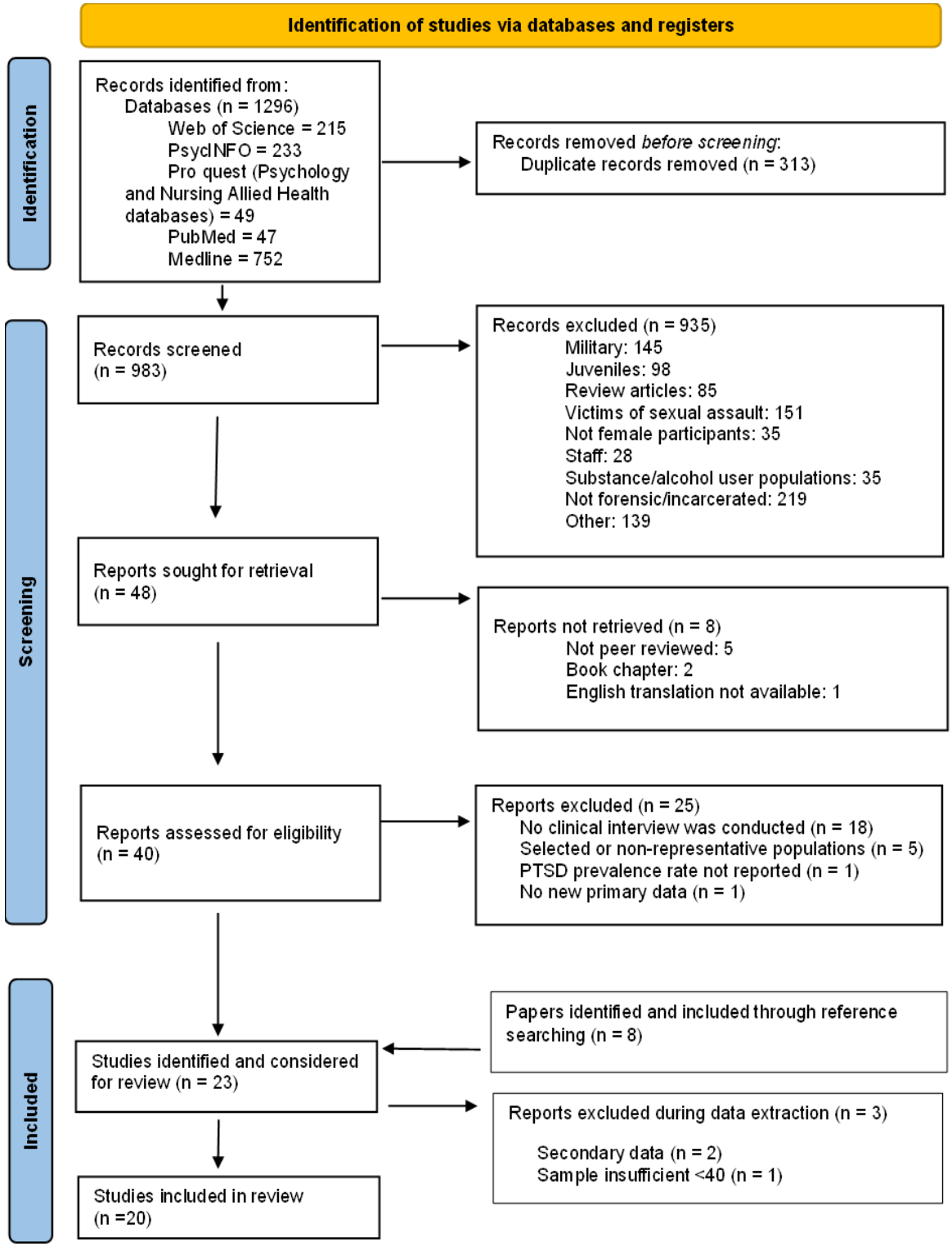
Exclusion criteria		Justification
1.	Reviews and commentaries	Reviews and commentaries were excluded as they do not contain primary data
2.	Studies published only in languages other than English	Studies published only in languages other than English were excluded due to lack of time and resources to translate these articles
3.	Studies that select participants from one specific population of prisoners (e.g. military)	Prevalence across the general prison population was desired to ascertain the PTSD prevalence in this demographic. Therefore, other populations such as community or military samples were not included
4.	Studies that used self-report screening measures to determine if PTSD is present	Self-report screening measures are not included as The National Institute for Health and Care Excellence (NICE) guidelines indicate that a clinical interview is required to determine PTSD diagnosis (NICE, 2018).
5.	Studies that include children or adolescents	Prevalence across female adult prison population was sought, and so studies including children or adolescents were excluded
6.	Clinical Trials	Clinical trials were not included as exclusion criteria have already been applied in these studies, and these samples are therefore not representative of the general prison population
7.	Studies with less than 40 participants	Studies with less than 40 participants were excluded to avoid computational difficulties of handling zero prevalence rates in

		studies where the sample was likely to have been selected for reasons other than the calculation of prevalence.
8.	Studies that included participants who were not incarcerated	Prevalence across the general prison population was desired to ascertain the PTSD prevalence in this demographic. Therefore, other populations such as community samples were excluded
9.	Studies that include only male participants	Prevalence across female adult prison population was sought, and so studies only including male participants were excluded
10.	Studies that only included lifetime prevalence of PTSD	Lifetime prevalence of PTSD was excluded, as we aimed to determine the current prevalence of PTSD within incarcerated populations

The initial search identified a total of 1296 articles, which was reduced to 983 once duplicates were removed. The results of the application of the inclusion and exclusion criteria to the results of the search are presented in Figure 1.1. The most common reasons for exclusion were: using non incarcerated populations (n=219), juvenile populations (n=98), or using populations from a specific subset, such as victims of sexual assault (n=151), and military personnel (n=145). The full text of 48 articles were then sought for retrieval, and full articles were reviewed in more depth against the exclusion criteria. 8 additional articles were also identified from the references of the included studies. Twenty articles satisfied the criteria for inclusion, and thus will be included in this meta-analysis.

Figure 1.1

Results of the systematic search and the application of the inclusion criteria



The 20 studies included in the meta-analysis were from 11 different countries, with a total of 7266 participants. The countries included were: Australia (Butler et al., 2005; Heffernan et al., 2015; Tye & Mullen, 2006), Brazil (Andreoli et al., 2014), Chile (Mundt et al., 2016), China (Huang et al., 2006; Zhong et al., 2020), Ecuador (Molina-Coloma et al., 2022), French Guiana (Nacher et al., 2018), Germany (Mir et al., 2015), Ireland (Mohan et al., 1997), New Zealand (Brinded et al., 2001), South Africa (Naidoo et al., 2022), USA (Combs et al., 2019; Gunter et al., 2008; Konecky & Lynch, 2019; Lynch et al., 2014; Teplin et al, 1996; Trestman et al., 2007; Zlotnick et al., 1997). Sample sizes varied from 45 – 2703 participants and included both sentenced and remand incarcerated populations. Details of all included studies can be found in Table 1.3.

Table 1.3

Details of included studies

Author, year	Country	Sample size	Legal status	Sampling method	Diagnostic instrument	Interviewers
Andreoli et al. (2014)	Brazil	617	Unknown	Random stratified sample	CIDI	Trained interviewers
Brinded et al. (2001)	New Zealand	162	Sentenced and remand	All females were invited	CIDI	Trained interviewers, clinical psychology students and psychiatric nurses
Butler et al. (2005)	Australia	176	Sentenced and remand	Random stratified sample	CIDI (modified for national survey)	Psychology masters students and nurses
Combs et al. (2019)	USA	83	Sentenced and remand	Random	CAAPE-5	Trained interviewers
Gunter et al. (2008)	USA	56	Sentenced	Random	MINI+	Trained interviewers
Heffernan et al. (2015)	Australia	65	Sentenced and remand	All females were invited	CIDI	Indigenous researchers with mental health experience
Huang et al. (2006)	China	471	Sentenced	Random	CAPS	Psychiatrists
Konecky and Lynch (2019)	USA	152	Unknown	Random online number generator	CAPS-5	Unknown
Lynch et al. (2014)	USA	491	Sentenced and remand	Random	CIDI	Trained interviewers
Mir et al. (2015)	Germany	150	Sentenced and remand	All females were invited	MINI	Clinical psychologist
Mohan et al. (1997)	Ireland	45	Sentenced and remand	Random	SCAN	Psychiatrist trainees
Molina-Coloma et al. (2022)	Ecuador	50	Unknown	Unknown	MINI	Unknown
Mundt et al. (2016)	Chile	198	Sentenced and remand	All females were invited	MINI	Clinical psychologist and nurses
Nacher et al. (2018)	French Guiana	60	Unknown	All females were invited	MINI	Psychiatrists and nurses
Naidoo et al. (2022)	South Africa	126	Sentenced and remand	Random stratified sample	SCID-5	Forensic psychiatrist
Teplin et al. (1996)	USA	1272	Remand	Random stratified sample	NIMH DIS III-R	Trained interviewers
Trestman et al. (2007)	USA	201	Sentenced and remand	Random	SCID-II and CAPS	Research staff; students and professionals
Tye and Mullen (2006)	Australia	103	Sentenced and remand	Midnight census – all eligible invited	Adapted CIDI	Forensic mental health professionals
Zhong et al. (2020)	China	2703	Sentenced	Cluster sampling	MINI	Psychiatrists
Zlotnick (1997)	USA	85	Sentenced	Random	SCID	Unknown

Data extraction

All data was extracted by the author. It is expected that event rates in the included papers will typically be reported as the number of participants who do and do not meet the diagnostic criteria for PTSD. These event rates will be extracted from each study. Where event rates are not reported directly, then they could be estimated from percentages, or reported regressions (e.g. logistic regressions). However, regression-based event rates are often calculated from data that has been adjusted for the association with one or more additional covariates. These adjustments highlight the idiosyncratic character of the reported regression coefficients, and therefore may result in dissimilarity with the effects reported within the other primary studies. The contribution of standardised regression coefficients to overall heterogeneity will be explored empirically if high levels of heterogeneity are identified using the random effects model.

There may be multiple reporting of event rates within a single study if the authors report several measures of PTSD, or report the same outcome measure in different population subgroups. Wherever possible, multiple outcomes will be combined in a single quantitative outcome using the procedures described by Borenstein et al. (2009). However, where this is not possible, then any multiple effects will be directly included in the meta-analysis. If multiple reporting of outcomes from that same primary study are included, then this could result in a slight reduction in confidence intervals for the random effects model as the sample size of that particular study will be counted for each reported effect.

Defining problematic variance

Study level effects may be found to be heterogeneous if they present with variation from the meta-analytic average that cannot be attributed to the true variation in the typical distribution of PTSD in this population. Heterogeneity may occur due to variation in the method design, measurement error, or other factors that have not been adequately controlled.

Higgins I^2 is a commonly used measure of heterogeneity and will be used in this meta-analysis. Higher values of I^2 indicate the presence of variation extraneous to the actual prevalence of PTSD in the population (e.g. variation that might be attributed to participant characteristics, methodological variation etc.). There is often considerable variation of methodology in primary studies for meta-analytic synthesis. Therefore, following the guideline for identifying high levels of heterogeneity described by Higgins, problematic heterogeneity was considered for Higgins I^2 values greater than 75% (Higgins et al., 2003). If problematic heterogeneity is found, then the subsequent analyses will focus on identifying the sources of heterogeneity between the estimates of PTSD in the primary studies, rather than on interpretation of the meta-analytic average.

Risk of Bias Assessment

To assess study level risk of bias, a list of quality criteria has been created. The quality criteria are based upon established risk of bias frameworks in the literature, including The Cochrane Collaboration Risk of Bias Tool (Higgins et al., 2011) and the Risk of Bias Assessment Tool for Nonrandomised Studies (Kim et al., 2013). The adapted framework assesses for risk of bias in six areas: selection bias, performance bias, detection bias, statistical bias, reporting bias and generalisation. Within each study, each area is assessed and categorised as low,

unclear, or high risk. The operationalisation of these areas of risk of bias is described in Table 1.4.

Table 1.4

Criteria for low, unclear and high risk of bias

Domain	Details	Risk of Bias
Selection Bias	Were efforts made to minimise selection bias? Was convenience sampling used? If so studies should potentially be penalised.	High risk - Within-subject design. Retrospective data collection. Convenience sampling with additional bias. Unclear risk - Convenience sampling without additional bias. Low risk - The study participants were consecutively recruited, or randomly sampled.
Performance Bias	Performance bias refers to systematic differences between/within groups in the participants motivation to complete the study.	High risk - Responses are not confidential or anonymous. Participants were rewarded for their participation in the study. Participants were told what questionnaires they were completing and why and any proposed hypotheses. Unclear risk - The study does not report levels of confidentiality and anonymity. It is not clear if participants were rewarded for their participation (e.g. motivation to respond in a certain way). It is unclear how much information was provided to the participant prior to taking part in the study Low risk - Study reports level of confidentiality and anonymity. Participants were not rewarded for their participation in the study Information and procedures are provided in a way that does not differentially motivated participants
Detection Bias	Is the design of the study is optimised to detect the effect in question? Ratings of design bias therefore reflect the position of the study design within the hierarchy of possible designs, with less optimal designs receiving a	High risk - The outcome measures were implemented differently across participants.

	penalty.	The outcome measures used had poor reliability and validity reported e.g. Cronbach's Alpha < 0.6. and/or test/retest reliability < 0.6 States that it has been translated but does not detail how this was conducted or clear problems in translation. Only using one dimension/ subscale of the scale or separating the subscales/ dimensions in the analysis. Unclear risk - Information regarding the outcome measures are either not reported or not clearly reported e.g. definition, validity, reliability. Cronbach's Alpha for outcome measures is between 0.6 and 0.7. Test-retest reliability for outcome measures is between .6 and .7 It is not clear if the measure was implemented consistently across all participants. Low risk - The outcome measures are clearly defined, valid and reliable, and are implemented consistently across all participants.
Statistical Bias	Have appropriate statistical methods been used? Is there incomplete data due to attrition?	High risk - Completer only analysis. Greater than 20% attrition. Unclear risk - Completer only analysis. Less than 20% attrition. Low risk - Appropriate statistical methods used. Less than 20% attrition.
Reporting Bias	Is there evidence of selective outcome reporting? i.e. only significant results reported. Are there measures that have not been reported in the results that have been mentioned in the method section?	High risk - Not reported full outcome measures that are stated in the method section/reported only a subsample of results/only significant results. Unclear risk - Not all descriptive and/or summary statistics are presented. Low risk - Reported all results of measures as outlined in the method.
Generalisation	Can the research findings be applied to settings other than that in which they were originally tested? Are there any differences between the study participants and those persons to whom the review is applicable?	High risk - Small sample with or without idiosyncratic features Sample below 40 = excluded Sample of 40 - 59 = High risk of bias The sample size is not adequate to detect an effect. Unclear risk - Sufficient sample for generalisation but with some idiosyncratic features Sample = 60 - 99).

	Low risk - Sufficient sample for generalisation and representative of target population Sample >100).

Note Domains of risk of bias and the criteria for ratings of low, unclear or high risk are presented

The application of these risks of bias, and the allocation of low, unclear or high risk to the individual studies is described in Table 1.5.

Table 1.5

Bias ratings for each of the included studies

Study name	Selection Bias	Performance Bias	Detection Bias	Statistical Bias	Reporting Bias	Generalisability	Overall Quality Index
Andreoli et al. (2014)	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	48%
Brinded et al. (2001)	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	48%
Butler et al. (2005)	Unclear risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	45%
Combs et al. (2019)	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	45%
Gunter et al. (2008)	Unclear risk	High risk	Low risk	Low risk	Low risk	Unclear risk	43%
Heffernan et al. (2015)	Low risk	Low risk	Unclear risk	Low risk	Low risk	Unclear risk	48%
Huang et al. (2006)	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	52%
Konecky and Lynch (2019)	Low risk	High risk	Low risk	Unclear risk	Low risk	Low risk	43%
Lynch et al. (2014)	Low risk	High risk	Unclear risk	Low risk	Low risk	Low risk	45%
Mir et al. (2015)	Low risk	Unclear risk	Low risk	Low risk	Low risk	Low risk	50%
Mohan et al. (1997)	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	High risk	43%
Molina-Coloma et al. (2021)	Unclear risk	Unclear risk	High risk	Unclear risk	High risk	Unclear risk	33%
Mundt et al. (2016)	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	48%

Nacher et al. (2018)	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	Unclear risk	45%
Naidoo et al. (2022)	Unclear risk	Low risk	Low risk	Low risk	Low risk	Low risk	50%
Teplin et al. (1996)	Low risk	High risk	Low risk	Low risk	Low risk	Low risk	48%
Trestman et al. (2007)	Unclear risk	Low risk	Low risk	Low risk	Low risk	Low risk	50%
Tye and Mullen (2006)	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	71%
Zhong et al. (2020)	Unclear risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	45%
Zlotnick (1997)	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	Unclear risk	45%

Note Ratings of risk of bias. Red indicates high risk of bias, amber marks an unclear risk of bias and green is a low risk of bias. An overall quality index has also been calculated based on risk of bias, and is presented as a percentage score.

Selection Bias

Overall, selection bias was low within the studies. Twelve studies were rated as low risk of bias with seven rated as unclear risk of bias. The studies that were rated as low-risk outlined a clear recruitment process (Andreoli et al., 2014; Brinded et al., 2001; Heffernan et al., 2015; Huang et al., 2006; Konecky and Lynch, 2019; Lynch et al., 2014; Mir et al., 2015; Mohan et al., 1997; Mundt et al., 2016; Nacher et al., 2018; Teplin et al., 1996; Zlotnick, 1997). The studies rated as unclear risk of bias were often vague when describing their recruitment process or appeared to use sampling procedures that may be subject to bias (e.g. convenience sampling) (Butler et al., 2005; Combs et al., 2019; Gunter et al., 2008; Molina-Coloma et al., 2021; Naidoo et al., 2022; Trestman et al., 2007; Tye and Mullen, 2006; Zhong et al., 2020).

Performance Bias

Performance bias was mixed within the studies, with a large number of the studies rated as unclear or high risk of bias. The majority of studies did not describe whether confidentiality and anonymity were discussed and did not disclose whether participants were rewarded for their participation. Due to the bias that may occur from these issues, these studies were rated

as having an unclear risk of performance bias. Four studies were rated as having a high risk due to the use of incentive-based recruitment (Gunter et al., 2008; Konecky and Lynch, 2019; Lynch et al., 2014; Teplin et al., 1996). Five studies were classed as low risk of bias as confidentiality and anonymity were both discussed, and participants were not compensated for taking part in the study (Heffernan et al., 2015; Huang et al., 2006; Naidoo et al., 2022; Trestman et al., 2007; Tye and Mullen., 2006).

Detection Bias

Eight papers were rated as low risk of detection bias, as outcome measures were clearly defined, valid and reliable, and are implemented consistently across all participants. Eleven papers were rated as unclear bias (Andreoli et al., 2014; Brinded et al., 2001; Butler et al., 2005; Heffernan et al., 2015; Lynch et al., 2014; Mohan et al., 1997; Mundt et al., 2016; Nacher et al., 2018; Tye and Mullen, 2006; Zhong et al., 2020; Zlotnick, 1997) as basic psychometric properties of the included measures were not reported. One paper was rated as high risk (Molina-Coloma et al., 2021) as it used only the PTSD section of the MINI diagnostic which could be indicative of a high risk of confirmation bias.

Statistical Bias

Twenty papers were rated as low risk for this area of bias, with two papers rated as unclear. One study was rated as unclear due to not reporting an attrition rate (Molina-Coloma et al., 2021). One study was rated as unclear risk of bias due to using the average score of subscales in place of missing data (Konecky and Lynch, 2019).

Reporting Bias

Overall, the full reporting of the outcomes within the studies was considered to be good, with nineteen papers being classed as low risk of reporting bias as there was no evidence of selective outcome reporting. The papers reported all results of measures as outlined in the method. One paper was rated high risk (Molina-Coloma et al., 2021) as they reported in their methodology that an outcome measure had been used, and the results of this were not reported.

Generalisability

Fourteen studies were rated as low risk for generalisability bias. Small sample sizes contributed to generalisability bias, as the sample may not be large enough to detect prevalence of PTSD. Six studies being rated as unclear risk due to having between 60 – 99 participants in the sample (Combs et al., 2019; Gunter et al., 2008; Heffernan et al., 2015; Molina-Coloma et al., 2021; Nacher et al., 2018). One study was rated as high risk due to having a sample size less than 60. The results obtained from these studies should therefore be interpreted with caution, as these samples may be too small to detect prevalence of PTSD and so may skew the data.

Summary

Overall, there was a mixed level of bias across the studies included in the meta-analysis, indicating that results may need to be interpreted with caution. Only one included study did not report any unclear or high risk of bias in any of the quality criteria, (Huang et al., 2006). However, fourteen studies did not report any high risk of bias in any area, which is positive (Andreoli et al., 2014; Brinded et al., 2011; Butler et al., 2005; Combs et al., 2019; Heffernan et al., 2015; Mir et al., 2015; Mundt et al., 2016; Nacher et al., 2018; Naidoo et al., 2022;

Trestman et al., 2007; Tye and Mullen, 2006; Zhong et al., 202; Zlotnick, 1997). The highest area for risk of bias was performance bias, due to confidentiality and anonymity not being discussed, or because participants were compensated for taking part in research. Due to the low number of studies in this field, studies with medium to high risk of bias were included. Consequently, the results of this meta-analysis should be interpreted with caution. However, the studies included are felt to be a representative summary of the research literature as it currently stands, and it is hoped that future research will adopt better methodological standards.

Results

Selection of the meta-analytic model

The fixed effects model weights the average of the study level effects proportional to the size of the studies sample. Therefore, the fixed-effect model is of use if all the studies included in the meta-analysis share a single, high quality, methodology. However, due to high number of uncontrolled factors found in psychological studies and the variety of methodologies that employed, effect sizes are very likely to vary (e.g., as a reflection of methodological weakness across studies, uncontrolled moderators, natural variation in the effect that is being measured). If the researcher has collected data from a variety of studies that have different methodological strengths and weaknesses, it is unlikely that all the studies are functionally equivalent in this way. It is expected that there will be differences in the participants, methods and measurement, and this would have affected the estimate of prevalence. Consequently, it is

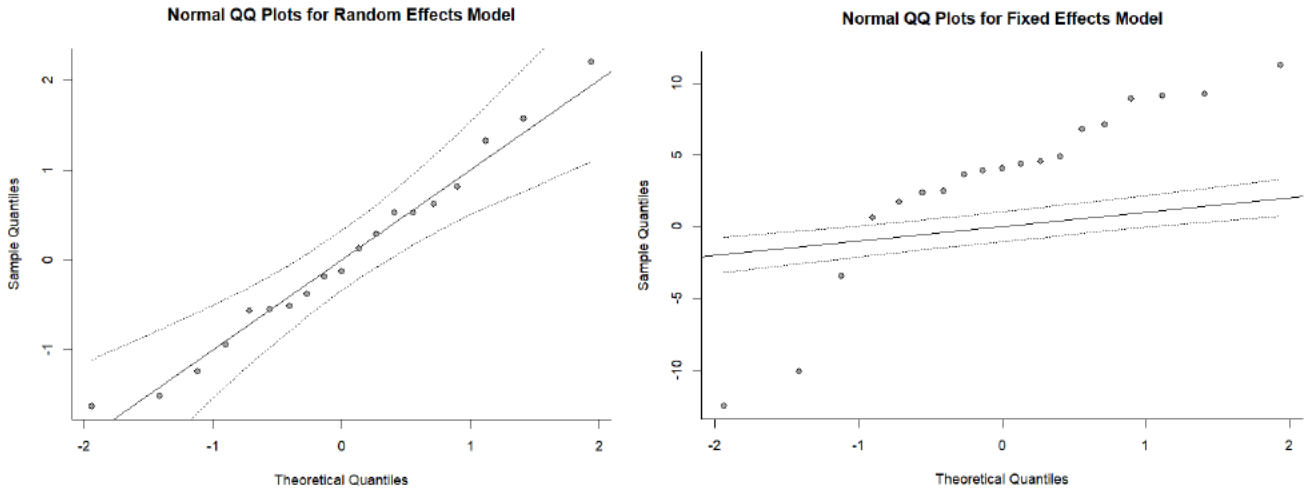
not appropriate to assume a common effect size. In these cases, the random-effects model is more suitable than the fixed-effect model.

The Random Effects Model (REM) weights the average of the effects by the sample size, and the weighting is also inversely related to how discrepant it is from the meta-analytic average. The REM therefore includes study level information about between studies variation, and precision of measurement.

The distributions of primary study effects are shown below in Figure 1.2 using both the fixed effects model and the random effects model. The between studies variance (tau^2) was calculated using the Restricted Maximum Likelihood estimator (REML). This estimator was used as there is research to suggest it is more robust to deviations from normality (Banks et al., 1985).

Figure 1.2

QQ plots for Random Effects and Fixed Effects Models



Note QQ plots for Random Effects Model, and Fixed Effects Model. Each primary study is plotted within the QQ plots, the 95% confidence interval is also included in this plot. Studies conforming to normality and linearity assumptions will fall within the 95% confidence interval.

When using the fixed effect model there is clear evidence of non-normality in the distribution of study level effects. However, as can be seen from Figure 1.2, there is no evidence of non-normality in the distribution of PTSD within the primary studies when using the Restricted Maximum Likelihood estimator (REML). Therefore, this indicates that the use of the random effects model using the REML estimate is an appropriate method for coping with between studies variation in the calculation of the weighted average effect.

The omnibus test

There were 20 studies reporting a total of 7266 participants. Participants were selected from general adult female prison populations, using inclusion and exclusion criteria defined in Table 1.2. Table 1.6 provides a summary of the prevalence rates reported in each of the included studies.

Prevalence was calculated in different ways using a variety of clinical measures. Typically point prevalence was calculated from cross-sectional studies and reflected current PTSD morbidity at the time of testing. Period prevalence was calculated over different time periods, ranging from 6-month prevalence rates (Teplin et al., 1996), to 12-month prevalence rates (Andreoli et al., 2014; Butler et al., 2005; Heffernan et al., 2015; Lynch et al., 2014; and Teplin et al., 1996).

Table 1.6

Prevalence rates of PTSD in each of the included studies

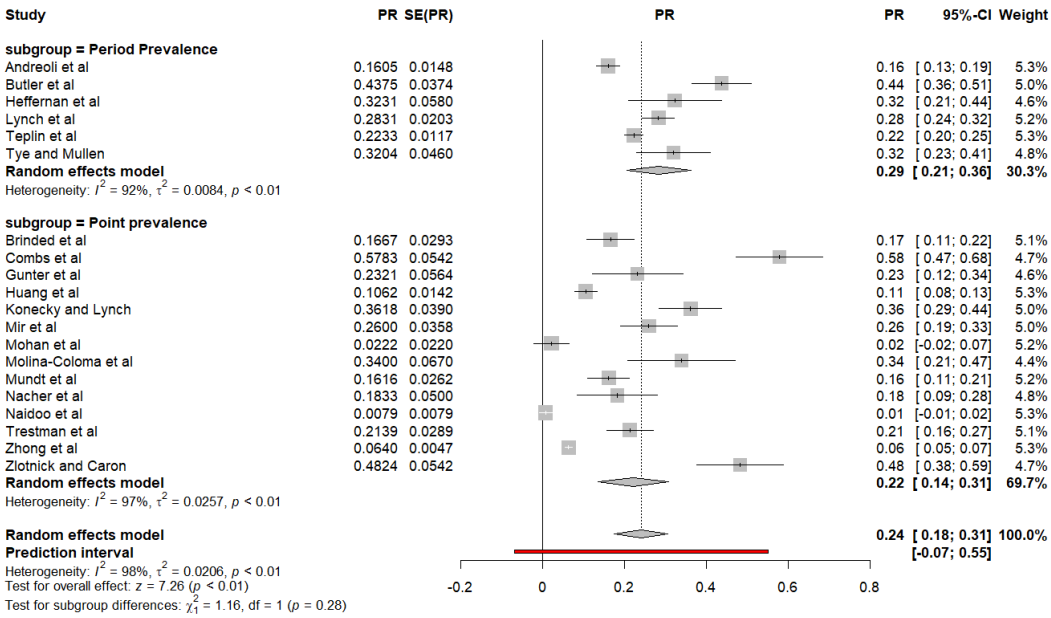
Study	Effect	Prevalence type	Std. Error	C.I. Lower	C.I. Upper	Weight(random)
Andreoli et al. (2014)	0.160454	Period	0.014776	0.131494	0.18941	47.95
Brinded et al. (2001)	0.166667	Point	0.02928	0.109278	0.22406	46.53
Butler et al. (2005)	0.4375	Period	0.037393	0.36421	0.51079	45.38
Combs et al. (2019)	0.578313	Point	0.054205	0.472074	0.68455	42.42
Gunter et al. (2008)	0.232143	Point	0.056419	0.121564	0.34272	41.98
Heffernan et al. (2015)	0.323077	Period	0.058005	0.209389	0.43676	41.67
Huang et al. (2006)	0.106157	Point	0.014194	0.078338	0.13398	47.99
Konecky and Lynch (2019)	0.361842	Point	0.038976	0.28545	0.43823	45.14
Lynch et al. (2014)	0.283096	Period	0.020331	0.243248	0.32294	47.51
Mir et al. (2015)	0.26	Point	0.035814	0.189805	0.33019	45.62
Mohan et al. (1997)	0.022222	Point	0.021974	-0.020846	0.06529	47.35
Molina-Coloma et al. (2021)	0.34	Point	0.066993	0.208697	0.47130	39.80
Mundt et al. (2016)	0.161616	Point	0.02616	0.110344	0.21289	46.90
Nacher et al. (2018)	0.183333	Point	0.049954	0.085426	0.28124	43.23
Naidoo et al. (2022)	0.007937	Point	0.007905	-0.007557	0.02343	48.31
Teplin et al. (1996)	0.22327	Period	0.011676	0.200385	0.24616	48.14
Trestman et al. (2007)	0.21393	Point	0.028925	0.157239	0.27062	46.57
Tye and Mullen (2006)	0.320388	Point	0.045978	0.230273	0.41050	43.96
Zhong et al. (2020)	0.064003	Point	0.004708	0.054776	0.07323	42.42
Zlotnick (1997)	0.482353	Point	0.054199	0.376125	0.58858	42.41

As can be seen in Figure 1.3, for the period prevalence estimate, the REM suggested a weighted average of prevalence of PTSD of 0.29 and a 95% confidence interval of between 0.21 to 0.36. For the point prevalence estimates the random effects model suggested a weighted average of prevalence of PTSD was only marginally lower, at 0.22 (95% confidence interval = 0.14 to 0.31). There was no statistically significant difference between the period prevalence and point prevalence estimates ($x2 = 1.16$, $p = 0.28$). When period prevalence and

point prevalence estimates were combined, the weighted average prevalence rate was 0.24 with a 95% confidence interval of between 0.18 to 0.31. A forest plot of PTSD prevalence divided into studies that report point prevalence and those that report period prevalence (inclusive of 6 and 12 month prevalence rates) is depicted in Figure 1.3.

Figure 1.3

Forest plot of period and point PTSD prevalence.



As can be seen in the Figure 1.3, a high level of heterogeneity in the primary studies was observed (Higgin's I^2 = 98%, tau2 = 0.0206, p < 0.01), suggesting that the estimates of PTSD in the primary studies may be biased by the presence of uncontrolled or confounding factors. It should also be noted that high estimates of heterogeneity were observed for both the point prevalence (I^2 = 97%) and period prevalence estimates (I^2 = 92%). Therefore, the focus

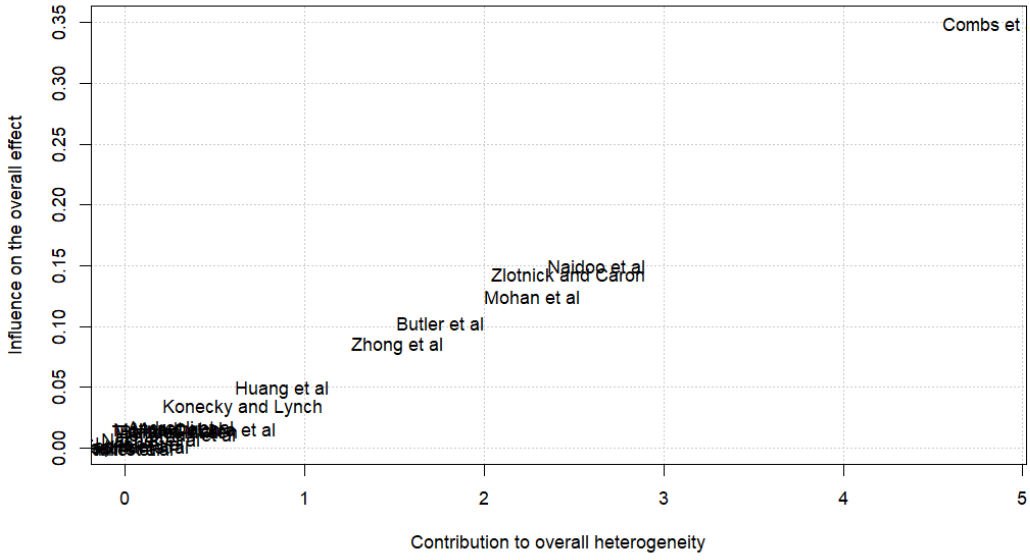
of the subsequent analyses will be upon the identification of the sources of heterogeneity between the estimates of PTSD in the primary studies. As no significant difference was observed between different ways of calculating prevalence rates, these data were pooled for subsequent sensitivity analyses.

The impact of influential primary studies

The impact of influential studies was analysed using a 'leave one out' method. The random effects model was calculated removing each of the primary studies one at a time and the change in weighted average effect size and change in heterogeneity was recorded for the omission of each study level effect. This, therefore, gives us an indication of the influence of each of the primary studies. The results of this 'leave one out' analysis are presented in a Baujat plot in Figure 1.4 (Baujat et al., 2002).

Figure 1.4

Baujat diagnostic plot of sources of heterogeneity



Note. The vertical axis reports the influence of the study on the overall effect, and the horizontal axis reports the discrepancy of the study with the rest of the literature.

The Baujat chart indicates that the removal of Combs et al. (2019) from the meta-analysis had the most impact on the estimated effect and the measure of between studies heterogeneity. The REM was recalculated with Combs et al. (2019) removed. The corrected REM reported a weighted average prevalence of 0.22 (95% CI 0.16 to 0.28). The corrected REM evidences an approximately 8% decrease relative to the uncorrected estimate. Accordingly, the corrected estimate did not change the substantive conclusions of this synthesis.

The effect of risk of bias in the primary studies

To assess the impact of study level risk of bias upon heterogeneity, subgroup analysis has been conducted on the risk of bias ratings of "low risk" and "any risk". For the purpose of this analysis, unclear and high risk of bias have been combined. This analysis has been conducted for each of the six types of methodological bias.

Table 1.7

Subgroup analysis of effect sizes and 95% Confidence Intervals for low risk and any risk (which includes unclear and high risk of bias)

	Low Risk			Any Risk				
	Effect	**95% CI**	**k**	**Effect**	**95% CI**	**k**	**X^2**	**P**
Selection bias	0.2224	(0.158 to 0.2901)	12	0.2697	(0.1381 to 0.4013)	8	0.39	0.5315
Performance bias	0.1872	(0.0665 to 0.3079)	5	0.2590	(0.1819 to 0.3360)	15	0.97	0.3258
Detection bias	0.2434	(0.1275 to 0.3593)	8	0.2397	(0.1589 to 0.3204)	12	0.00	0.9592
Statistical bias	0.2297	(0.1600 to 0.2994)	18	0.3563	(0.2903 to 0.4223)	2	6.68	0.0098
Reporting bias	0.2297	(0.1600 to 0.2994)	18	0.3563	(0.2903 to 0.4223)	2	5.97	0.0145
Generalisability bias	0.2090	(0.1437 to 0.2742)	13	0.3055	(0.1642 to 0.4468)	7	1.48	0.2243

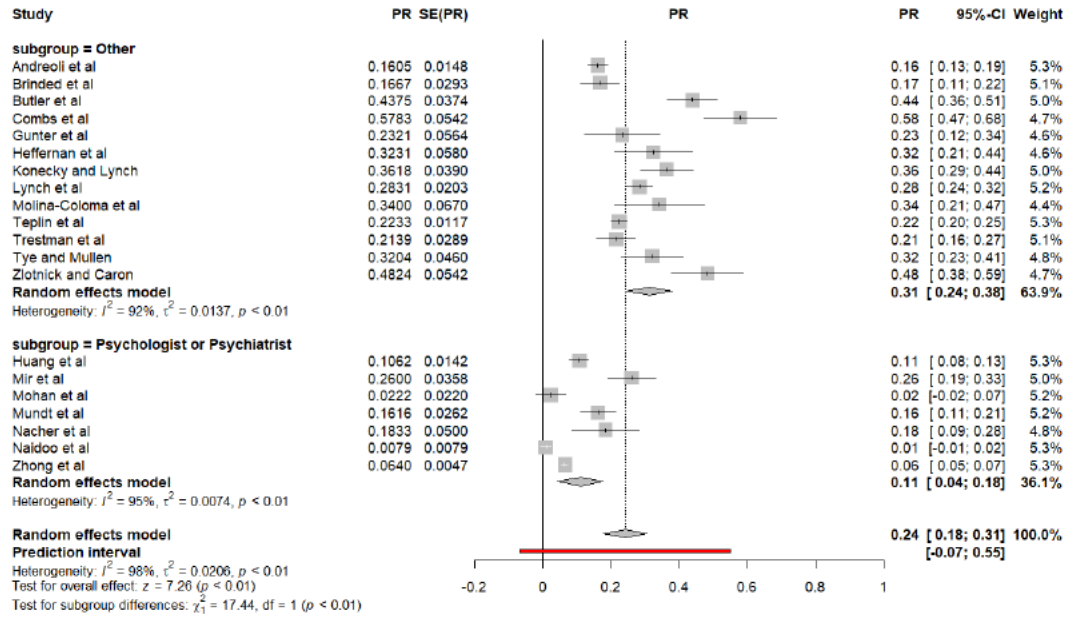
As can be seen in Table 1.7, statistical bias and reporting bias evidenced statistically significant differences in prevalence of PTSD, with lower levels of bias being associated with lower estimates of prevalence. This suggests that inclusion of studies that are at risk of statistical bias and reporting bias may inflate the estimate of the prevalence of PTSD. There are no statistically significant differences between levels of bias in estimates of prevalence for selection, performance, detection and generalisability biases.

The impact of rater expertise on the estimate of prevalence

In order to estimate the impact of rater expertise on the estimate of prevalence, a subgroup analysis was conducted between psychology and psychiatry qualified raters or trainees, and 'other' raters, which included, trained interviewers, and masters level students. Where psychology and psychiatry raters were mixed with other raters, they were categorised in the psychology and psychiatry category. The outcome of this comparison is shown in Figure 1.5.

Figure 1.5

Forest plot of rater experience and estimate of PTSD prevalence including 95% confidence intervals



Note. Subgroups are divided into rater experience, with prevalence estimates of psychologists and psychiatrists, versus 'other' category, which includes students, trained interviewers, and other mental health professionals with unknown experience.

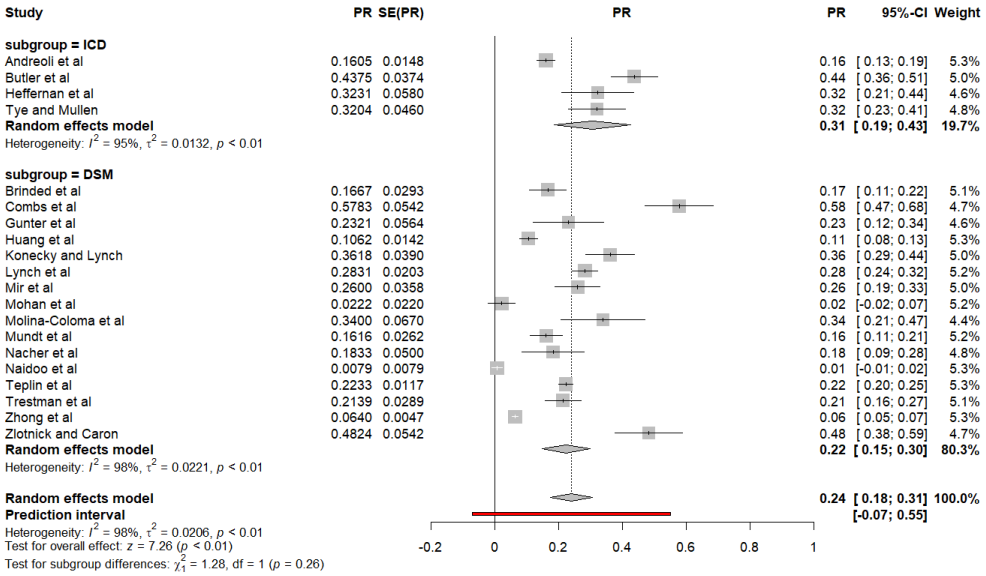
As can be seen from the forest plot in Figure 1.5, there is a significant difference between raters ($X^2 = 17.44$, $p<0.01$), with the estimates from psychologists and psychiatrists reporting lower estimates of PTSD prevalence. This suggests that the use of raters who are likely to be less familiar with psychiatric diagnosis, tend to inflate the estimate of the prevalence of PTSD.

The impact of diagnostic criteria on the estimate of prevalence

In order to estimate the impact of diagnostic criteria on the estimate of prevalence, a subgroup analysis was conducted between ICD and DSM diagnostic criteria. The outcome of this comparison is shown in Figure 1.6.

Figure 1.6

Forest plot of PTSD prevalence rates and 95% Confidence Intervals, comparing studies that used ICD and DSM diagnostic criteria



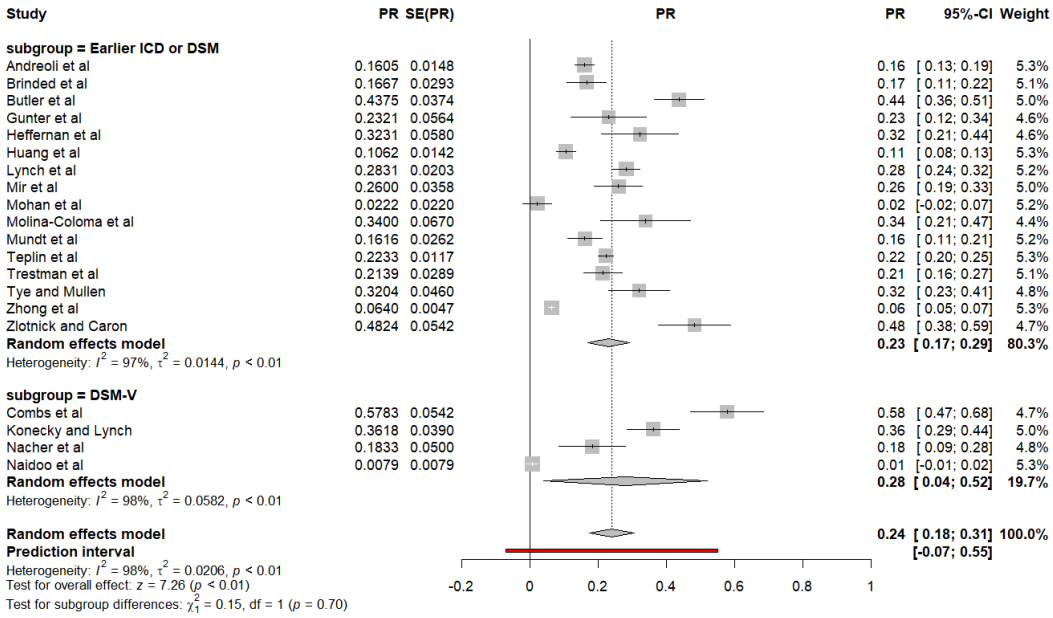
As can be seen from the forest plot in Figure 1.6, there is no significant difference ($X^2 = 1.28$, $p<0.26$), between estimates of prevalence when using the ICD and DSM diagnostic criteria.

The impact of new diagnostic criteria in DSM-5 on the estimate of prevalence

In the DSM-5, the PTSD criteria were updated, with changes to the existing criteria, and the addition of new symptoms. Therefore, a subgroup analysis was conducted to examine whether there is a statistical difference between estimates of prevalence from studies who used the DSM-5, and those who used any other criterion, including DSM-III, DSM-IV, and ICD-10. No papers used the ICD-11 criteria. The outcome of this comparison is shown in Figure 1.7.

Figure 1.7

Forest plot of DSM-5 and other diagnostic criteria, and estimate of prevalence including 95% Confidence Intervals



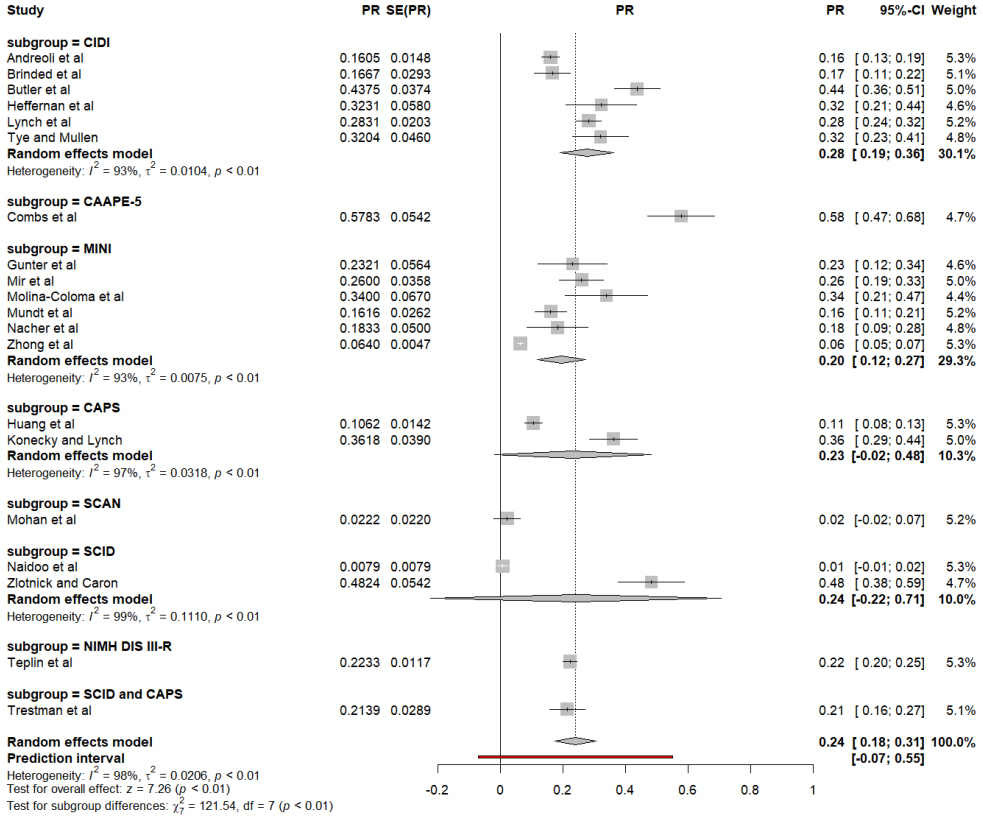
As can be seen from the forest plot in Figure 1.7, there is no significant difference ($X^2 = .15$, $p<0.70$), between estimates of prevalence when using the DSM-5, compared with other diagnostic criteria such as DSM-III, DSM-IV or ICD-10.

The impact of the diagnostic instrument used on the estimate of prevalence

A variety of different interview based diagnostic instruments were used to determine a diagnosis of PTSD within this meta-analysis. They included the following: Composite International Diagnostic Instrument (CIDI), The Mini International Neuropsychiatric Interview (MINI), Clinician administered PTSD Scale for DSM-5 (CAPS-5), Structured Clinical Interview for DSM-5 (SCID-5), Schedule for Clinical Assessment in Neuropsychiatry (SCAN), Comprehensive Addictions and Psychological Evaluation – 5 (CAAPE-5), and National Institute of Mental Health Diagnostic Interview Schedule Version III-R (NIMH DIS-III-R). Therefore, a subgroup analysis was conducted to examine whether there is a statistical difference between estimates of prevalence from studies using different diagnostic instruments. The outcome of this comparison is shown in Figure 1.8.

Figure 1.8

Forest plot of estimate of prevalence and diagnostic instrument including 95% Confidence Intervals



Note. Forest plot of estimate of PTSD prevalence, comparing studies that used different diagnostic instruments to determine PTSD diagnosis, including CIDI, CAAPE-5, MINI, CAPS, SCAN, SCID and NIMH DIS III-R.

As can be seen from the forest plot in Figure 1.8, there is a significant difference (X^2 = 121.54, p<0.01), between estimates of prevalence when using the different diagnostic instruments. However, it should be noted that is determined by a single study (Combs et.al., 2019) which used the CAAPE-5, and reported a significantly higher prevalence rate than other studies. This was the only study to use this diagnostic instrument.

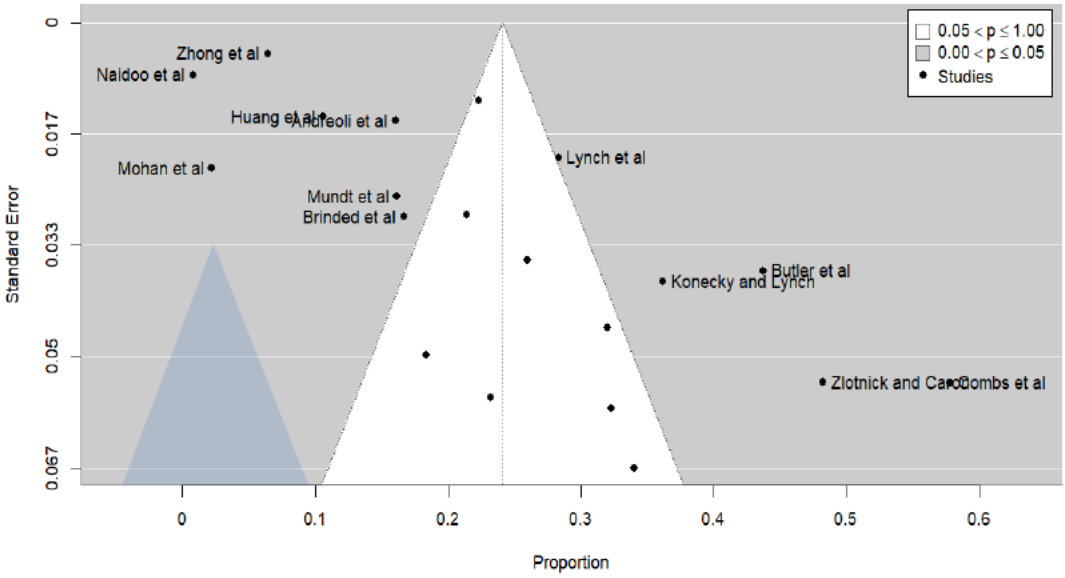
The impact of publication and small study biases

Publication bias occurs due to the propensity for statistically significant results to be published over non-significant results. This leads to the so called 'file drawer' problem, where the failure to publish non-significant results leads to a false impression of significance within the existent literature. Small study bias is when studies with smaller sample sizes show larger variability in their measurements of prevalence. A preponderance of small studies within the literature can result in an inflation of perceived heterogeneity.

Both publication bias and small study bias can be seen in a funnel plot, which plots the magnitude of the study's prevalence rate (i.e., the importance of the study in the meta-analysis) and estimates how much the study deviates from the meta-analytic average. When publication bias is not present, studies with small sample sizes will scatter more widely at the bottom of the plot, compared to studies that have larger samples and their effect size is more likely to be closer to the meta-analytic average. This should create a symmetrical triangular (of funnel) shape. If studies with small sample sizes and non-significant results are not prevalent in the funnel plot, then it is indicative of publication bias. This publication bias is likely to lead to an over-estimation of the true prevalence of PTSD. The funnel plot of prevalence rates in the included studies is included in Figure 1.9.

Figure 1.9

Funnel plot of the prevalence of PTSD



Note. The 95% confidence interval of the expected distribution of prevalence is shown as an inverted white "funnel". The area of the blue triangle depicts the area of the plot associated with smaller sample sizes and prevalence rates consistent with the normal population.

As can be seen from Figure 1.9, previously mentioned high level of heterogeneity is clearly visible, with a large number of studies falling outside the 95% confidence interval for the weighted average effect. However, there is also some evidence of publication bias in the distribution of study level prevalence rates, in that there appears to be a lack of studies in the area of the funnel plot associated with small sample size reporting prevalence rates that are consistent with that of the general population (as depicted as a blue triangle in Figure 1.9). This conclusion is confirmed by a significant Eggers test of funnel graph asymmetry (t=4.70, p <.01).

A trim and fill procedure (Duval & Tweedle, 2000) was used to attempt to simulate the effect of publication bias. This procedure works on the assumption that publication bias leads to a

funnel plot that is not symmetrical. The procedure therefore removes the most extreme small studies from the side of the funnel plot associated with positive effects, and recalculates the effect size at each iteration until the funnel plot is symmetrical about the (corrected) effect size. Although the trim and fill procedure achieves an adjusted effect size, it also reduces the variance in effect sizes and so causes biased and smaller confidence intervals. Consequently, the original studies are returned into the analysis, and the trim and fill procedure calculates a mirror image for each study on the side of the funnel plot associated with negative effects. Unfortunately, the trim and fill algorithm failed to impute any additional studies and thus could not provide a corrected random effects model. An alternative would be to calculate Orwin's failsafe number (Orwin, 1983). This method calculates the number of studies with non-significant results (i.e., reporting prevalence at general population levels) which would need to be included in the meta-analysis for the overall effect to be reduce to a minimally interpretable prevalence rate that might be consider different to that of the general population. This procedure suggests that 58 studies reporting an average prevalence of 5% would be required to reduce the observed prevalence rate of 24% to a value of 10% and 3799 studies would be required to reduce the observed prevalence rate of 24% to a value of 5%. Given, that these failsafe numbers are substantially larger than the 20 studies that were included in the meta-analysis, it can be concluded that the difference between the general population prevalence of PTSD and that observed in a female prison population is unlikely to be merely an artifact of studies missing due to publication bias.

Comparison with previous studies

A previous meta-analysis by Baranyi et al. (2018) reported a point prevalence of 21.1% (95% CI: 16.9, 25.6) in a female prison population. Our current estimate of point prevalence of

PTSD in a female prison population is 22% (95% CI: 31, 14) which is consistent ($X2 < 0.03$, $p = 0.85$) with that provided by Baranyi et al. (2018). Baranyi et al., also reported a 12-month prevalence rate of 26.1% (95% CI: 15.9, 37.8). Our estimate was for period prevalence rates, containing both 6 and 12- month prevalence rates, and was found to be 29% (95% CI: 36, 21). This is consistent ($X2 < 0.56$, $p = 0.45$) with that provided by Baranyi et al. (2018).

Discussion

This meta-analysis was based upon 20 samples, consisting of 7266 participants from 11 countries worldwide. However please note that no studies from the UK were included in this meta-analysis. Both point prevalence and period prevalence (consisting of 6 and 12-month prevalence rates) were established. Point prevalence ranged from 6.4 – 57.8%, and period prevalence ranged from 16 – 43.7%. Average pooled point prevalence was found to be 22% and pooled period prevalence was found to be 29%. There were no statistical differences found between average pooled point and period prevalence rates. When combined, there was an average prevalence rate of 24%. This research is consistent with a previous meta-analysis from 2018, which found a point prevalence rate of PTSD in female prison populations to be 21.1%, and a 12-month prevalence of 26%. This can be compared to a prevalence rate of 6.2% found in male prison populations (Baranyi et al., 2018), and between 0.3 – 1.9% for point prevalence in the general population, and 1.5 – 3.6% for 12- month prevalence (Koenen et al., 2017). This strengthens the literature surrounding the high prevalence of PTSD internationally in female prison populations, and implications for forensic services need to be considered.

A high level of heterogeneity was found in the literature, which may have contributed to the variance in prevalence rates. There is a large variety of diagnostic tools used to assess for PTSD, and there were statistically significant differences found between prevalence rates from different diagnostic tools. However, this was largely driven by one study (Combs et al., 2019) which was the only study to use the CAAPE-5 diagnostic instrument. The CAAPE-5 has previously been found to demonstrate good reliability, and a 95% agreement with the SCID which is hailed as gold standard of diagnostic interviews (Gallagher et al., 2006; Proctor & Hoffermann, 2012). Combs et al. (2019) noted that their study, conducted at a rural detention facility, might yield a higher PTSD prevalence compared to other studies. They proposed that this disparity could be attributed to the inclusion of a larger proportion of individuals on remand, who are awaiting sentencing. Consequently, their sample includes individuals who might transition to a community release program or psychiatric facility, who may not be included in prison prevalence studies.

There were also significant differences found between prevalence rates depending on the profession of staff conducting the interviews, with lower prevalence rates found when psychiatrists or psychologists complete the diagnostic interview, compared with those completed by trained interviewers, or other mental health professionals. When conducted by psychiatrists or clinical psychologists, the prevalence rate was found to be 11%, compared to 31% when conducted by other mental health professionals or trained interviewers. It is possible that the use of raters who are likely to be less familiar with psychiatric diagnosis, tend to inflate the estimate of the prevalence of PTSD. The accuracy of prevalence rates may therefore be increased by ensuring that participants are assessed by an appropriately trained clinician. There were no significant differences found between estimates of prevalence rates

when using the DSM-5, compared to the DSM III, DSM IV or ICD-10. None of the studies included used ICD-11 criteria, and so congruence between DSM-5 and ICD-11 could not be assessed.

There are some limitations of this meta-analysis to consider. Some of the heterogeneity in this analysis can be accounted for by the sensitivity analyses conducted, such as differences in measurement tools and interviewers. However, it should be noted that there may be additional sources of heterogeneity within the literature that have not been accounted for. Sampling methods and differences between sub types of populations, such as those on remand or sentenced, were not investigated in this meta-analysis due to small sample sizes. These differences may account for some of the additional variance between studies. As well as high levels of heterogeneity, the research was also found to be of varying quality, with a significant risk of bias found in some of the literature. Statistically significant differences in prevalence of PTSD were found between research with varying risk of bias, with lower levels of statistical and reporting bias being associated with lower estimates of prevalence. This suggests that inclusion of studies that are at risk of statistical bias and reporting bias may inflate the estimate of the prevalence of PTSD. It is important that future research focuses on improving methodological quality to reduce risk of bias. This can be achieved by following the 'low risk' guidelines as outlined in Table 1.4, which are based on The Cochrane Collaboration Risk of Bias Tool (Higgins et al., 2011) and the Risk of Bias Assessment Tool for Nonrandomised Studies (Kim et al., 2013). An additional limitation of this review is that no studies from the UK were included in the meta-analysis, and so caution should be taken when applying the findings of this review to the UK. The only studies in the UK concerning prevalence of PTSD in female prison populations that were found in the literature search used

self-report measures rather than clinical interviews, and so were not included in this review due to exclusion criteria concerning risk of bias from self-report measures. It is suggested that additional research should be conducted in the UK to establish the prevalence of PTSD in female prison populations using clinical interviews.

As previously highlighted, the female prison population is expanding at a staggering rate, and has increased by nearly 60% since the year 2000 (Fair & Walmsley, 2022). Considering that the prevalence of PTSD in prison populations is significantly higher in females than males (Baranyi et al., 2018), the difficulty of supporting people in prisons with PTSD is an increasing problem that forensic services are facing. PTSD is often undiagnosed and untreated in prison populations, and is associated with higher rates of re-offending, suicidality, self-harm, aggressive behaviour, substance abuse and co-morbid mental health conditions which cause significant problems for the prison service (Kubiak, 2004; Karatzias et al., 2018; Facer-Irwin et al., 2019). There is evidence to indicate that an individual is also less likely to benefit from rehabilitative programmes in prison if they are experiencing PTSD (Allely & Allely, 2020). There is therefore an increasing need for the prison service to focus on the needs of those with PTSD in the prison system.

Increasing rates of diagnosis and treatment of PTSD in prisons could substantially improve outcomes for individuals with PTSD. There is significant evidence for the positive impact of PTSD treatment in the community, including for therapies such as eye movement desensitisation and reprocessing (EMDR) and trauma-focussed cognitive behavioural therapy (TF-CBT) which are outlined in NICE guidance for PTSD (Mavranezouli et al., 2020a; NICE, 2018; Watts et al., 2013). Research indicates that after interventions, there are significant

reductions in trauma symptoms (Mavranezouli et al., 2020a; Watts et al., 2013). PTSD is also associated with functional impairments, increased mortality rates and poor quality of life, and treatment of PTSD such as EMDR and TF-CBT have been found to be a cost-effective in the community due to improvements in functional impairment, cost to mental health services and health complications associated with PTSD (Mavranezouli et al., 2020b). In prison populations, a recent meta-analysis of 16 trauma focussed interventions found a small, but significant effect sizes for reductions in trauma symptoms after treatment (Malik et al. 2023). The meta-analysis found significant reductions in trauma symptoms for both individual and group interventions, as well as for interventions that used stabilisation, and those that used memory reprocessing techniques. The interventions included in the meta-analysis were prison-based programmes such as Strive and Thrive, Beyond Trauma and Seeking Safety. Larger effect sizes were found for interventions delivered individually, and those that included memory reprocessing techniques. However, Malik (2023) highlights that the evidence base in prison populations is still somewhat limited, and calls for further research to be conducted to improve the standard of trauma treatment offered in prison settings. Notably, the two most widely recommended and evidenced based interventions, EMDR and TF-CBT, were absent from the review (NICE, 2018; Malik, 2023). It is therefore recommended that future research in prisons includes EMDR and TF-CBT interventions. As well as evaluating the impact on trauma symptoms, it would be helpful if future research including EMDR and TF-CBT can also investigate the effect on recidivism rates, so the wider impact of these interventions can be assessed.

There is also evidence to suggest that a focus on trauma informed care, not just trauma focussed interventions, may be beneficial. Covington (2003) defines trauma-informed

services as ‘services that have been created to provide assistance for problems other than trauma, but in which all practitioners have a shared knowledge base and/or core understanding about trauma resulting from violence’. Research with female offenders found that trauma-informed planning could improve safety in prisons, including reductions in disciplinary adjudications and conflict between inmates (Benedict, 2014). Trauma informed environments aim to enhance feelings of safety and security (King, 2017), and this can be a significant challenge in prison environments. Prisons are highly reactive environments, where confinement to cells, lack of personal space and agency, physical violence, use of authority and difficult interpersonal relationships are commonplace (Auty et al., 2023). Prison life may also echo past traumas for women, and simulate previous abuse dynamics, leading to ongoing re-traumatisation (Dirks, 2004). In the UK, the Ministry of Justice released a Female Offender Strategy which outlined its ambition to implement a trauma informed approach in all women’s prisons (Ministry of Justice, 2018). The proposal includes an aim to continue trauma awareness training for prison staff, expand the availability of a prisoner-led trauma support group, a commitment to adopt gender informed practices and environments, and consider alternatives to custody such as electronic monitoring in the community (Ministry of Justice, 2018). As well as improvements in prison safety, there is also evidence to suggest that trauma informed care could reduce recidivism. A longitudinal study by Lehrer (2001) found strong evidence of reduced recidivism rates for incarcerated women who took part in trauma-informed programmes compared to those who did not which adds further support to the value to trauma informed care in women’s prisons. Recent evidence has also supported the use of gender-responsive and trauma-specific brief interventions. Messina & Schepps (2021) found that interventions entitled ‘Healing Trauma for Women’ and ‘Exploring Trauma for Men’ found a significant reduction in mental health symptomology and aggression in people who

attended the courses. Additionally, they found that the effect was most pronounced for those who had experienced the most ACEs, indicating that these interventions may be most effective with people who have been exposed to the high levels of trauma.

There are several implications to consider based on this review. The high prevalence of trauma experienced in female prison populations serves as a reminder that women are often victims first, before offending. Research indicates that women experience disproportionately high rates of trauma, and early interventions to mitigate traumatic experiences should be urgently considered and may prevent future offending (Allely & Allely, 2000; Farrington & Welsh, 2008). Secondly, trauma informed approaches and appropriate treatments should be provided in custodial settings, with evidence to suggest that the use of gender-responsive interventions may be helpful (Messina et al, 2021). Despite evidence supporting the beneficial effects of PTSD treatment in prisons, it has been estimated that approximately 90% of incarcerated individuals with PTSD are not receiving the appropriate services needed for their mental health (Jacobowitz et al., 2017). It is important to note that the prison system is legally responsible for medical care for inmates, and must provide appropriate treatment, including mental health services (Miller & Najavits, 2012) and so this highlights the large need for an increase in services. Improvements in rates of recognition of PTSD, diagnosis and treatment in forensic services are required (Prost et al., 2022). Finally, as summarised out by the Ministry of Justice's (2018) Female Offender Strategy, substantial adaptations to the justice system are vital, to accommodate women's needs and the disproportionate mental health inequalities. Research indicates that trauma interventions and trauma informed care may improve both the quality of life of offenders, the safety in prisons and reductions in reoffending (Benedict, 2014; Leher, 2001). Incarceration has the potential unique opportunity

to intervene and treat people who have experienced substantial trauma and improve both personal and social outcomes substantially (Butler et al., 2005; Sindicich et al., 2014). It is imperative that appropriate adaptations are made to the way women are supported in the justice system, and evidence-based treatments are offered to reduce the disastrous impact for these women who have experienced such high levels of trauma, and the consequent detrimental impacts for their families and society.

References

Allely, C. S., & Allely, B. (2020). Post traumatic stress disorder in incarcerated populations: Current clinical considerations and recommendations. *Journal of Criminal Psychology, 10*(1), 30–42. https://doi.org/10.1108/JCP-09-2019-0044

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV* (4. ed.). doi:10.1176/appi.books.9780890423349

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Andreoli, S. B., Dos Santos, M. M., Quintana, M. I., Ribeiro, W. S., Blay, S. L., Taborda, J. G. V., & de Jesus Mari, J. (2014). Prevalence of mental disorders among prisoners in the state of Sao Paulo, Brazil. *PloS one*, *9*(2), e88836. https://doi.org/10.1371/journal.pone.0088836

Armytage, P., Martyres, K., & Feiner, M. (2000). Women in corrections: Getting the balance right. In Women in Corrections: Staff and Clients Conference convened by the Australian Institute of Criminology in conjunction with the Department for Correctional Services, South Australia, Adelaide, South Australia (Vol. 31).

Auty, K. M., Liebling, A., Schliehe, A., & Crewe, B. (2023). What is trauma-informed practice? Towards operationalisation of the concept in two prisons for women. Criminology & Criminal Justice, 23(5), 716-738. https://doi.org/10.1177/17488958221094980

Banks, B. D., Mao, I. L., & Walter, J. P. (1985). Robustness of the Restricted Maximum Likelihood Estimator Derived Under Normality as Applied to Data with Skewed

Distributions1. Journal of Dairy Science, 68(7), 1785–1792. https://doi.org/10.3168/jds.S0022-0302(85)81028-6

Baranyi, G., Cassidy, M., Fazel, S., Priebe, S., & Mundt, A. P. (2018). Prevalence of posttraumatic stress disorder in prisoners. *Epidemiologic reviews*, *40*(1), 134-145. https://doi.org/10.1093/epirev/mxx015

Baujat, B., Mahe, C., Pignon, J.-P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. Statistics in Medicine, 21(18), 2641–2652. https://doi.org/10.1002/sim.1221

Bisson, J. I., Ehlers, A., Matthews, R., Pilling, S., Richards, D., & Turner, S. (2007). Psychological treatments for chronic post-traumatic stress disorder: Systematic review and meta-analysis. *The British Journal of Psychiatry*, *190*(2), 97–104. https://doi.org/10.1192/bjp.bp.106.021402

Borenstein, M. (Ed.). (2009). Complex Data Structures. In Introduction to meta-analysis. Chichester, U.K: John Wiley & Sons.

Bowler, N., Phillips, C., & Rees, P. (2018). The association between imported factors and prisoners' mental health: Implications for adaptation and intervention. *International Journal of Law and Psychiatry*, *57*, 61–66. https://doi.org/10.1016/j.ijlp.2018.01.001

Brinded, P. M., Simpson, A. I., Laidlaw, T. M., Fairley, N., & Malcolm, F. (2001). Prevalence of psychiatric disorders in New Zealand prisons: a national study. *Australian & New Zealand Journal of Psychiatry*, *35*(2), 166-173. https://doi.org/10.1046/j.1440-1614.2001.00885.x

Butler, T., Allnutt, S., Cain, D., Owens, D., & Muller, C. (2005). Mental disorder in the New South Wales prisoner population. *Australian & New Zealand Journal of Psychiatry*, *39*(5), 407-413. https://doi.org/10.1080/j.1440-1614.2005.01589.x

Byrne, M., & Howells, K. (2000). Key issues in the provision of correctional services of women. Paper presented at the Women in Corrections: Staff and Clients Conference, Australian Institute of Criminology, Adelaide, SA. https://ro.uow.edu.au/cgi/viewcontent.cgi?article=3660&context=sspapers

Campbell, C. A., Albert, I., Jarrett, M., Byrne, M., Roberts, A., Phillip, P., Huddy, V., & Valmaggia, L. (2016). Treating Multiple Incident Post-Traumatic Stress Disorder (PTSD) in an Inner City London Prison: The Need for an Evidence Base. *Behavioural and Cognitive Psychotherapy*, *44*(1), 112–117. https://doi.org/10.1017/S135246581500003X

Cavadino, M., & Dignan, J. (2006). Penal policy and political economy. Criminology & Criminal Justice, 6(4), 435-456. https://doi.org/10.1177/1748895806068581

Clark, V. A., & Duwe, G. (2024). Sex Differences in the Effects of Adverse Childhood Experiences on Institutional Misconduct among Adults in Prison.

Cloud, D. H., Parsons, J., & Delany-Brumsey, A. (2014). Addressing mass incarceration: A clarion call for public health. *American Journal of Public Health*, 104(3), 389-391. https://doi.org/10.2105/AJPH.2013.301741

Combs, E., Guston, K., Kopak, A., Raggio, A., & Hoffmann, N. G. (2019). Posttraumatic stress, panic disorder, violence, and recidivism among local jail detainees. *International Journal of Prisoner Health*, *15*(4), 366-375. https://doi.org/10.1108/IJPH-06-2018-0036

Covington, S. (2003). Beyond trauma. *Center City, MN: Hazeldon*.

Covington, S. (2018). Women in prison: Approaches in the treatment of our most invisible population. In *Breaking The Rules* (pp. 141-155). Routledge.

Coyle, A., & Fair, H. (2016). Imprisonment worldwide: the current situation and an alternative future. Policy Press.

Dalsklev, M., Cunningham, T., Dempster, M., & Hanna, D. (2021). Childhood Physical and Sexual Abuse as a Predictor of Reoffending: A Systematic Review. *Trauma, Violence, & Abuse*, *22*(3), 605–618. https://doi.org/10.1177/1524838019869082

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Dierkhising, C. B. (2014). *Snare or Turning Point? An Exploration of Change and Continuity in Criminal Behavior Among Formerly Incarcerated Youth*. University of California, Riverside. https://escholarship.org/uc/item/3k91r013

Dirks, D. (2004). Sexual Revictimization and Retraumatization of Women in Prison. *Women's Studies Quarterly*, *32*(3/4), 102–115. https://www.jstor.org/stable/40004582

Doi, S. A. R., & Thalib, L. (2008). A Quality-Effects Model for Meta-Analysis: *Epidemiology*, *19*(1), 94–100. https://doi.org/10.1097/EDE.0b013e31815c24e7

Dulisse, B. C., Logan, M. W., & Hazelwood, A. R. (2023). Evaluating the use of PTSD Assessment Tools Within the U.S. Prison Setting: An Exploratory Analysis. *The Prison Journal*, *103*(3), 307–328. https://doi.org/10.1177/00328855231173144

Duval, S. & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Facer-Irwin, E., Blackwood, N. J., Bird, A., Dickson, H., McGlade, D., Alves-Costa, F., & MacManus, D. (2019). PTSD in prison settings: A systematic review and meta-analysis of

comorbid mental disorders and problematic behaviours. *PLOS ONE*, *14*(9), e0222407. https://doi.org/10.1371/journal.pone.0222407

Facer-Irwin, E., Karatzias, T., Bird, A., Blackwood, N., & MacManus, D. (2022). PTSD and complex PTSD in sentenced male prisoners in the UK: Prevalence, trauma antecedents, and psychiatric comorbidities. *Psychological Medicine*, *52*(13), 2794–2804. https://doi.org/10.1017/S0033291720004936

Fair, H., & Walmsley, R. (2022). World female imprisonment list. ICPR. https://www.icpr.org.uk/news-events/2022/world-female-prison-population-60-2000

Farrington, D. P., & Welsh, B. C. (2008). Saving children from a life of crime: Early risk factors and effective interventions.

Firmin, M. W., Pugh, K. C., Turkelson, L., Annecharico, K. N., & Sohn, V. A. (2015). Implications for patients regarding PTSD diagnostic criteria differences when using the DSM and ICD classification systems. *International Journal of Psychology Research*, *10*(3), 247. https://www.proquest.com/scholarly-journals/implications-patients-regarding-ptsd-diagnostic/docview/1842831546/se-2

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). Structured Clinical Interview for DSM-5: Research Version. Arlington, VA: American Psychiatric Association. https://apsychoserver.psychofizz.psych.arizona.edu/JJBAReprints/PSYC621/694A/SCID/SCID-5-CV.pdf

Fleming, J., Levy, M., Kaldor, J., Donovan, B., & Butler, T. (2001). Childhood sexual abuse among Australian prisoners. Venereology, 14(3). https://search.informit.org/doi/abs/10.3316/informit.558380917498634

Forbes, D., Creamer, M., & Biddle, D. (2001). The validity of the PTSD checklist as a measure of symptomatic change in combat-related PTSD. *Behaviour research and therapy*, *39*(8), 977-986. https://doi.org/10.1016/S0005-7967(00)00084-X

Forbes, D., Lockwood, E., Phelps, A., Wade, D., Creamer, M., Bryant, R. A., McFarlane, A., Silove, D., Rees, S., Chapman, C., Slade, T., Mills, K., Teesson, M., & O†Donnell, M. (2013). Trauma at the Hands of Another: Distinguishing PTSD Patterns Following Intimate and Nonintimate Interpersonal and Noninterpersonal Trauma in a Nationally Representative Sample. *The Journal of Clinical Psychiatry*, *74*(2), 21205. https://doi.org/10.4088/JCP.13m08374

Fovet, T., Wathelet, M., Amad, A., Horn, M., Belet, B., Roelandt, J.-L., Thomas, P., Vaiva, G., & D'Hondt, F. (2023). PTSD in prison settings: The need for direct comparisons with the general population. *Psychological Medicine*, *53*(2), 597–599. https://doi.org/10.1017/S0033291721000507

Friedman M. J. (2013). Finalizing PTSD in DSM-5: Getting Here From There and Where to Go Next. *Journal of Traumatic Stress*, *26*(5), 548–556. https://doi.org/10.1002/jts.21840

Galatzer-Levy, I. R., & Bryant, R. A. (2013). 636,120 Ways to Have Posttraumatic Stress Disorder. *Perspectives on Psychological Science*, *8*(6), 651–662. https://doi.org/10.1177/1745691613504115

Gallagher, S. M., Penn, P. E., Brooks, A. J., & Feldman, J. (2006). Comparing the CAAPE, A New Assessment Tool for Co-occurring Disorders, with the SCID. *Psychiatric Rehabilitation Journal*, *30*(1), 63–65. https://doi.org/10.2975/30.2006.63.65

Goldhill, R. (2009). The Corston report: Reading between the lines. *The Prison Service Journal*, *184*, 13-19.

Grella, C. E., Lovinger, K., & Warda, U. S. (2013). Relationships among trauma exposure, familial characteristics, and PTSD: A case-control study of women in prison and in the general population. *Women & Criminal Justice*, *23*(1), 63-79. https://doi.org/10.1080/08974454.2013.743376

Griffin, M. G., Uhlmansiek, M. H., Resick, P. A., & Mechanic, M. B. (2004). Comparison of the posttraumatic stress disorder scale versus the clinician-administered posttraumatic stress disorder scale in domestic violence survivors. *Journal of Traumatic Stress*, *17*(6), 497–503. https://doi.org/10.1007/s10960-004-5798-4

Gunter, T. D., Arndt, S., Wenman, G., Allen, J., Loveless, P., Sieleni, B., & Black, D. W. (2008). Frequency of mental and addictive disorders among 320 men and women entering the Iowa prison system: use of the MINI-Plus. *Journal of the American Academy of Psychiatry and the Law Online*, *36*(1), 27-34.

Hafstad, G. S., Thoresen, S., Wentzel-Larsen, T., Maercker, A., & Dyb, G. (2017). PTSD or not PTSD? Comparing the proposed ICD-11 and the DSM-5 PTSD criteria among young survivors of the 2011 Norway attacks and their parents. *Psychological Medicine*, *47*(7), 1283-1291. doi:10.1017/S0033291716002968

Heffernan, E., Andersen, K., Davidson, F., & Kinner, S. A. (2015). PTSD among Aboriginal and Torres Strait Islander people in custody in Australia: Prevalence and correlates. *Journal of Traumatic Stress*, *28*(6), 523-530. https://doi.org/10.1002/jts.22051

Heiskanen, M., & Lietonen, A. (2016). *Crime and gender: A study on how men and women are represented in international crime statistics*. Universität Tübingen. https://tobias-lib.ub.uni-tuebingen.de/xmlui/handle/10900/82974

Herman, J. L. (1992). Complex PTSD: A syndrome in survivors of prolonged and repeated trauma. *Journal of Traumatic Stress*, *5*(3), 377–391. https://doi.org/10.1002/jts.2490050305

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

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. BMJ, 327(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Huang, G., Zhang, Y., Momartin, S., Cao, Y., & Zhao, L. (2006). Prevalence and characteristics of trauma and posttraumatic stress disorder in female prisoners in China. *Comprehensive psychiatry*, *47*(1), 20-29. https://doi.org/10.1016/j.comppsych.2005.04.004

Hunt, J. C., Chesney, S. A., Jorgensen, T. D., Schumann, N. R., & deRoon-Cassini, T. A. (2018). Exploring the gold-standard: Evidence for a two-factor model of the Clinician Administered PTSD Scale for the DSM–5. *Psychological Trauma: Theory, Research, Practice, and Policy*, *10*(5), 551–558. https://doi.org/10.1037/tra0000310

Hyland, P., Shevlin, M., Fyvie, C., & Karatzias, T. (2018). Posttraumatic Stress Disorder and Complex Posttraumatic Stress Disorder in DSM-5 and ICD-11: Clinical and Behavioural Correlates. *Journal of Traumatic Stress*, *31*(2), 174–180. https://doi.org/10.1002/jts.22272

Jakobowitz, S., Bebbington, P., McKenzie, N., Iveson, R., Duffield, G., Kerr, M., & Killaspy, H. (2017). Assessing needs for psychiatric treatment in prisoners: 2. Met and unmet need. *Social Psychiatry and Psychiatric Epidemiology*, *52*(2), 231–240. https://doi.org/10.1007/s00127-016-1313-5

James, D., & Glaze, L. (2006). Mental Health Problems of Prison. *US Department of Justice, Bureau of Justice Statistics, Washington, DC.*

Karatzias, T., Power, K., Woolston, C., Apurva, P., Begley, A., Mirza, K., Conway, L., Quinn, C., Jowett, S., Howard, R., & Purdie, A. (2018). Multiple traumatic experiences, post-traumatic stress disorder and offending behaviour in female prisoners. *Criminal Behaviour and Mental Health*, *28*(1), 72–84. https://doi.org/10.1002/cbm.2043

Karl, A., Schaefer, M., Malta, L. S., Dörfel, D., Rohleder, N., & Werner, A. (2006). A meta-analysis of structural brain abnormalities in PTSD. *Neuroscience & Biobehavioural Reviews*, *30*(7), 1004–1031. https://doi.org/10.1016/j.neubiorev.2006.03.004

Karlsson, M. E., & Zielinski, M. J. (2020). Sexual Victimization and Mental Illness Prevalence Rates Among Incarcerated Women: A Literature Review. Trauma, Violence, & Abuse, 21(2), 326-349. https://doi.org/10.1177/1524838018767933

Katz, M. (2017). City of women: The old Melbourne gaol and a gender-specific interpretation of urban life. The Palgrave Handbook of Prison Tourism, 341-364. https://doi.org/10.1057/978-1-137-56135-0_17

Kerig, P. K., Mozley, M. M., & Mendez, L. (2020). Forensic assessment of PTSD via DSM-5 versus ICD-11 criteria: Implications for current practice and future research. Psychological Injury and Law, 13, 383-411. https://doi.org/10.1007/s12207-020-09397-4

Kim, S. Y., Park, J. E., Lee, Y. J., Seo, H.-J., Sheen, S.-S., Hahn, S., Jang, B.-H., & Son, H.-J. (2013). Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *Journal of Clinical Epidemiology*, *66*(4), 408–414. https://doi.org/10.1016/j.jclinepi.2012.09.016

King, E. A. (2017). Outcomes of Trauma-Informed Interventions for Incarcerated Women: A Review. *International Journal of Offender Therapy and Comparative Criminology*, *61*(6), 667–688. https://doi.org/10.1177/0306624X15603082

Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., Karam, E. G., Ruscio, A. M., Benjet, C., Scott, K., Atwoli, L., Petukhova, M., Lim, C. C. W., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Bunting, B., Ciutan, M., Girolamo, G., Degenhardt, L., Gureje, O., Haro, J. M., Huang, Y., Kawakami, N., Lee, S., Navarro-Mateu, F., Pennell, B. E., Pizza, M., Sampson, N., Have, M., Torres, Y., Viana, M. C., Williams, D., Xavier, M., & Kessler, R. C. (2017). Posttraumatic stress disorder in the World Mental Health Surveys. *Psychological Medicine*, *47*(13), 2260–2274. https://doi.org/10.1017/S0033291717000708

Konecky, E. M., & Lynch, S. M. (2019). Cumulative trauma exposure, emotion regulation, and posttraumatic stress disorder among incarcerated women. *Journal of Traumatic stress*, *32*(5), 806-811. https://doi.org/10.1002/jts.22435

Kubiak, S. P. (2004). The Effects of PTSD on Treatment Adherence, Drug Relapse, and Criminal Recidivism in a Sample of Incarcerated Men and Women. *Research on Social Work Practice*, *14*(6), 424–433. https://doi.org/10.1177/1049731504265837

Lehrer, D. (2021). Trauma-informed care: the importance of understanding the incarcerated women. *Journal of Correctional Health Care*, *27*(2), 121-126. https://doi.org/10.1089/jchc.20.07.0060

Levenson, J. S., & Willis, G. M. (2019). Implementing Trauma-Informed Care in Correctional Treatment and Supervision. *Journal of Aggression, Maltreatment & Trauma*, *28*(4), 481–501. https://doi.org/10.1080/10926771.2018.1531959

Lynch, S. M., DeHart, D. D., Belknap, J. E., Green, B. L., Dass-Brailsford, P., Johnson, K. A., & Whalley, E. (2014). A multisite study of the prevalence of serious mental illness, PTSD, and substance use disorders of women in jail. *Psychiatric services*, *65*(5), 670-674. https://doi.org/10.1176/appi.ps.201300172

MacDonald, M. (2013). Women prisoners, mental health, violence and abuse. *International Journal of Law and Psychiatry*, 36(3-4), 293-303. https://doi.org/10.1016/j.ijlp.2013.04.014

Malik, N., Facer-Irwin, E., Dickson, H., Bird, A., & MacManus, D. (2023). The Effectiveness of Trauma-Focused Interventions in Prison Settings: A Systematic Review and Meta-Analysis. *Trauma, Violence, & Abuse*, *24*(2), 844–857. https://doi.org/10.1177/15248380211043890

Maschi, T., Viola, D., Morgen, K., & Koskinen, L. (2015). Trauma, stress, grief, loss, and separation among older adults in prison: The protective role of coping resources on physical and mental well-being. *Journal of Crime and Justice*, *38*(1), 113-136. https://doi.org/10.1080/0735648X.2013.808853

Mavranezouli, I., Megnin-Viggars, O., Daly, C., Dias, S., Welton, N. J., Stockton, S., Bhutani, G., Grey, N., Leach, J., Greenberg, N., Katona, C., El-Leithy, S., & Pilling, S. (2020a). Psychological treatments for post-traumatic stress disorder in adults: A network meta-

analysis. *Psychological Medicine*, *50*(4), 542–555. https://doi.org/10.1017/S0033291720000070

Mavranezouli, I., Megnin-Viggars, O., Grey, N., Bhutani, G., Leach, J., Daly, C., Dias, S., Welton, N.J., Katona, C., El-Leithy, S., Greenburg, N., Stockton, S. & Pilling, S. (2020b). Cost-effectiveness of psychological treatments for post-traumatic stress disorder in adults. *PloS one*, *15*(4), e0232245. https://doi.org/10.1371/journal.pone.0232245

McCallum, K. (2018). Does PTSD predict institutional violence within a UK male prison population?. *Journal of Forensic Practice*, *20*(4), 229-238. https://doi.org/10.1108/JFP-12-2017-0055

McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD Checklist: A critical review. *Clinical Psychology Review*, *30*(8), 976–987. https://doi.org/10.1016/j.cpr.2010.06.012

Medlicott, D. (2012). Women in prison. In Handbook on prisons (pp. 275-297). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203118191-20/women-prison-diana-medlicott

Merrick, M. T., Ford, D. C., Ports, K. A., & Guinn, A. S. (2018). Prevalence of adverse childhood experiences from the 2011-2014 behavioral risk factor surveillance system in 23 states. *JAMA pediatrics*, *172*(11), 1038-1044. doi:10.1001/jamapediatrics.2018.2537

Messina, N., Grella, C., Burdon, W., & Prendergast, M. (2007). Childhood adverse events and current traumatic distress: A comparison of men and women prisoners. Criminal Justice and Behavior, 34, 1385–1401. https://doi.org/10.1177/0093854807305150

Messina, N. P., & Schepps, M. (2021). Opening the proverbial ‘Can of worms’ on trauma-specific treatment in prison: The association of adverse childhood experiences to treatment

outcomes. *Clinical Psychology & Psychotherapy*, *28*(5), 1210-1221. https://doi.org/10.1002/cpp.2568

Miller, N. A., & Najavits, L. M. (2012). Creating trauma-informed correctional care: A balance of goals and environment. *European Journal of Psychotraumatology*, *3*(1), 17246. https://doi.org/10.3402/ejpt.v3i0.17246

Ministry of Justice. (2018). Female offender strategy. London: HMSO. https://assets.publishing.service.gov.uk/media/5b3349c4e5274a55d7a54abe/female-offender-strategy.pdf

Mir, J., Kastner, S., Priebe, S., Konrad, N., Ströhle, A., & Mundt, A. P. (2015). Treating substance abuse is not enough: comorbidities in consecutively admitted female prisoners. *Addictive Behaviors*, *46*, 25-30. https://doi.org/10.1016/j.addbeh.2015.02.016

Mohan, D., Scully, P., Collins, C., & Smith, C. (1997). Psychiatric disorder in an Irish female prison. *Criminal Behaviour and Mental Health*, *7*(3), 229-235. https://doi.org/10.1002/cbm.175

Molina-Coloma, V., Salaberría, K., Pérez, J. I., & Kendall-Tackett, K. (2022). Traumatic events, psychological symptoms, and aggression in male and female prisoners. *Psychological trauma: theory, research, practice, and policy*, *14*(3), 480. https://doi.org/10.1037/tra0001039

Mundt, A. P., Kastner, S., Larraín, S., Fritsch, R., & Priebe, S. (2016). Prevalence of mental disorders at admission to the penal justice system in emerging countries: a study from Chile. *Epidemiology and psychiatric sciences*, *25*(5), 441-449. doi:10.1017/S2045796015000554

Nacher, M., Ayhan, G., Arnal, R., Basurko, C., Huber, F., Pastre, A., Jehel, L., Falissard, B. & About, V. (2018). High prevalence rates for multiple psychiatric conditions among inmates at French Guiana's correctional facility: diagnostic and demographic factors associated with violent offending and previous incarceration. *BMC psychiatry*, *18*, 1-9. https://doi.org/10.1186/s12888-018-1742-7

Naidoo, S., Subramaney, U., Paruk, S., & Ferreira, L. (2022). Mental illness and HIV amongst female inmates in Durban, South Africa. *South African Journal of Psychiatry*, *28*, 1628. https://journals.co.za/doi/full/10.4102/sajpsychiatry.v28i0.1628

National Institute for Health and Care Excellence. (2018). Post Traumatic Stress Disorder. https://www.nice.org.uk/guidance/NG116

Newson, J. J., Hunter, D., & Thiagarajan, T. C. (2020). The Heterogeneity of Mental Health Assessment. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.00076

O'Connor, M., Lasgaard, M., Spindler, H., & Elklit, A. (2007). The impact of different diagnostic criteria on PTSD prevalence. *Nordic Psychology*, *59*(4), 317–331. https://doi.org/10.1027/1901-2276.59.4.317

O'Donnell, M. L., Schaefer, I., Varker, T., Kartal, D., Forbes, D., Bryant, R. A. A., Silove, D., Creamer, M., McFarlane, A., Malhi, G., Felmingham, K., Van Hoof, M., Hadzi-Pavlovic, D., Nickerson, A., & Steel, Z. (2017). A systematic review of person-centered approaches to investigating patterns of trauma exposure. *Clinical Psychology Review*, *57*, 208–225. https://doi.org/10.1016/j.cpr.2017.08.009

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. Journal of Educational Statistics, 8, 157-159. https://doi.org/10.3102/10769986008002157

Pai, A., Suris, A. M., & North, C. S. (2017). Posttraumatic Stress Disorder in the DSM-5: Controversy, Change, and Conceptual Considerations. *Behavioral Sciences*, *7*(1), 7. https://doi.org/10.3390/bs7010007

Pettus-Davis, C., Renn, T., Lacasse, J. R., & Motley, R. (2019). Proposing a population-specific intervention approach to treat trauma among men during and after incarceration. *Psychology of Men & Masculinities*, *20*(3), 379–393. https://doi.org/10.1037/men0000171

Proctor, S. L., & Hoffmann, N. G. (2012). Identifying patterns of co-occurring substance use disorders and mental illness in a jail population. Addiction Research & Theory, 20(6), 492-503. https://doi.org/10.3109/16066359.2012.667853

Prost, S. G., Golder, S., Higgins, G. E., Pettus-Davis, C., Renn, T., Edmond, T., & Logan, T. K. (2022). Correlates of post-traumatic stress among victimized women on probation and parole. *Probation Journal*, *69*(1), 45-65. https://doi.org/10.1177/02645505211032171

Robles, R., Fresán, A., Evans, S. C., Lovell, A. M., Medina-Mora, M. E., Maj, M., & Reed, G. M. (2014). Problematic, absent and stigmatizing diagnoses in current mental disorders classifications: Results from the WHO-WPA and WHO-IUPsyS Global Surveys. International Journal of Clinical and Health Psychology, 14(3), 165-177. https://doi.org/10.1016/j.ijchp.2014.03.003

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Sadeh, N., & McNiel, D. E. (2015). Posttraumatic Stress Disorder Increases Risk of Criminal Recidivism Among Justice-Involved Persons With Mental Disorders. *Criminal Justice and Behavior*, *42*(6), 573–586. https://doi.org/10.1177/0093854814556880

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of clinical psychiatry*, *59*(20), 22-33. https://www.psychiatrist.com/wp-content/uploads/2021/02/15175_mini-international-neuropsychiatric-interview-mini.pdf

Sindicich, N., Mills, K. L., Barrett, E. L., Indig, D., Sunjic, S., Sannibale, C., Rosenfeld, J. & Najavits, L. M. (2014). Offenders as victims: post-traumatic stress disorder and substance use disorder among male prisoners. *The Journal of Forensic Psychiatry & Psychology*, *25*(1), 44-60. https://doi.org/10.1080/14789949.2013.877516

Summerfeldt, L. J., & Antony, M. M. (2002). Structured and semi structured diagnostic interviews. *Handbook of assessment and treatment planning for psychological disorders*, 3-37.

Teplin, L. A., Abram, K. M., & McClelland, G. M. (1996). Prevalence of psychiatric disorders among incarcerated women: I. Pretrial jail detainees. *Archives of general psychiatry*, *53*(6), 505-512. doi:10.1001/archpsyc.1996.01830060047007

Thompson, P. J., & Harm, N. J. (2000). Parenting from prison: Helping children and mothers. Issues in Comprehensive Paediatric Nursing, 23(2), 61-81. https://doi.org/10.1080/01460860050121402

Trestman, R. L., Ford, J., Zhang, W., & Wiesbrock, V. (2007). Current and lifetime psychiatric illness among inmates not identified as acutely mentally ill at intake in Connecticut's jails. *Journal of the American Academy of Psychiatry and the Law Online*, *35*(4), 490-500. https://pubmed.ncbi.nlm.nih.gov/18086741/

Tye, C. S., & Mullen, P. E. (2006). Mental disorders in female prisoners. *Australian & New Zealand Journal of Psychiatry*, *40*(3), 266-271. https://doi.org/10.1080/j.1440-1614.2006.01784.x

Tyler, N., Miles, H. L., Karadag, B., & Rogers, G. (2019). An updated picture of the mental health needs of male and female prisoners in the UK: Prevalence, comorbidity, and gender differences. *Social Psychiatry and Psychiatric Epidemiology*, *54*(9), 1143–1152. https://doi.org/10.1007/s00127-019-01690-1

Vitopoulos, N. A., Peterson-Badali, M., Brown, S., & Skilling, T. A. (2019). The Relationship Between Trauma, Recidivism Risk, and Reoffending in Male and Female Juvenile Offenders. *Journal of Child & Adolescent Trauma*, *12*(3), 351–364. https://doi.org/10.1007/s40653-018-0238-4

Walmsley, R. (2018). World prison population list (12th ed.). Institute for Criminal Policy Research.

Watts, B. V., Schnurr, P. P., Mayo, L., Young-Xu, Y., Weeks, W. B., & Friedman, M. J. (2013). Meta-Analysis of the Efficacy of Treatments for Posttraumatic Stress Disorder. *The Journal of Clinical Psychiatry*, *74*(06), e541–e550. https://doi.org/10.4088/JCP.12r08225

Weathers F. W., Blake D. D., Schnurr P. P., Kaloupek D. G., Marx B. P., Keane T. M. (2013). Clinician-Administered PTSD Scale for DSM-5 (CAPS-5). National Center for PTSD. https://doi.org/10.1037/pas0000486

World Health Organization (1990). Composite International Diagnostic Interview. World Health Organization; Geneva, Switzerland.

World Health Organization (2021). International statistical classification of diseases and related health problems (11th ed.). https://icd.who.int/

Zhong, S., & Wang, X. (2020). High psychiatric morbidity and comorbidity among female prisoners in Hunan, China. *Frontiers in Psychiatry*, *11*, 493977. https://doi.org/10.3389/fpsyt.2020.00271

Zlotnick, C. (1997). Posttraumatic Stress Disorder (PTSD), PTSD Comorbidity, and Childhood Abuse among Incarcerated Women. *The Journal of Nervous & Mental Disease, 185* (12), 761-763.

Chapter 2: An exploratory analysis of predictors of outcomes at HMP Grendon

Abstract

Introduction

HMP Grendon is the only prison in the UK that operates entirely as a Democratic Therapeutic Community. It has been called 'the jewel in the crown' of the British Prison Service due to its long-standing history of innovative rehabilitation, reduced reconviction rates and pro-social culture (Genders & Player, 1995; Marshall, 1997; Taylor, 2000). However, Grendon does not work for everyone, and little is known about who is likely to succeed, and who is not.

Method

A database of 3060 residents at HMP Grendon was collated, including psychometric data upon entry, and information about onwards move after leaving Grendon. Successful and non-successful outcomes were categorised based on meeting therapeutic goals, and progressive moves on from the prison. Using binary logistic regressions, a clinical regression decision tree was calculated to predict successful and non-successful outcomes.

Results

A clinical regression tree was able to predict non-successful outcomes at HMP Grendon with an overall accuracy of 97.6%, but successful outcomes only with a 7.9% accuracy. Men who

self-report high Antisocial Features, Treatment Rejection and Dominance, but low Social Self-Esteem, have a 98.1% chance of a non-successful outcome.

Discussion

This predictive model holds significant clinical implications, enabling the identification of treatment-interfering factors with a high level of confidence. This information could be used to adjust treatment accordingly and decrease the number of people that leave Grendon with a non-successful outcome. Unfortunately, the model is not as accurate at predicting successful outcomes. Further research is required to understand common psychometric profiles that predict successful outcomes.

Introduction

HMP Grendon in Buckinghamshire is a Category B men's prison with capacity for over 200 men. It opened in 1962 as an experimental psychiatric prison (Tollington, 1966), and was the first prison in the United Kingdom (UK) to operate as a Democratic Therapeutic Community (DTC). It has been described as 'the jewel in the crown' of the British Prison Service due to its long-standing history of innovative rehabilitation, reduced reconviction rates, pro-social culture, and humane treatment of individuals (Genders & Player, 1995; Marshall, 1997; Morris, 2004; Taylor, 2000). HMP Grendon accepts men who have committed serious violent and sexual offences, and are serving long determinate or indeterminate sentences. Men at Grendon are referred to as residents or members of the community, and so will be referred to in this way throughout this paper. At Grendon the traditional hierarchical structure found in prisons is replaced by a more liberal and participative culture (Brookes, 2010). To this day

Grendon is unique to the prison system, as it is the only facility in the UK where the whole prison operates as a DTC (Bennett & Shuker, 2017).

What is a Democratic Therapeutic Community?

Therapeutic communities are employed in parts of the UK prison system with the aim of rehabilitating prisoners, and are a radical alternative to mainstream prisons (Rawlings & Haigh, 2017). In the UK Category B male prison estate in England and Wales there are currently three DTCs: HMP Grendon, HMP Gartree and HMP Dovegate. There is also one DTC in the female prison estate, HMP Send.

DTCs form part of the Offender Personality Disorder pathway, which provides intervention to individuals with complex personality traits and forensic risk (Skett & Lewis, 2019). The main principles of a DTC are described in Cullen (1994):

- Responsibility – individual and collective responsibility are required. Each resident is asked to take responsibility for their own actions and discuss anything significant that may have happened.
- Empowerment – every member of the community has a direct say in how the community is run, including a democratically held right to vote other people out of therapy if they violate any of the main rules of the community (including no alcohol, drugs, sex, or violence).
- Support – the regime provides support to residents from a variety of staff, including psychologists, prison officers, probation staff, and educationalists – as well as from other residents who are trained to work in various support roles. This support is

considered important to provide a context for residents to speak openly and honestly about their problems.

- Confrontation – although it is important to provide an environment in which people feel safe, a therapeutic community is also a place where there is direct and candid confrontation of those who attempt to minimise their offending, or the harm that they have caused to their victims or others in the therapeutic community.

To attend Grendon, men voluntarily apply and are assessed against eligibility criteria (Bennett & Shuker, 2018). To be considered for a place at Grendon, men must be convicted and sentenced, with at least 18 months left on their sentence. They must have not been involved in violence, self-harm, or drug use for 6 months prior to applying, and they must not have any major current mental illness (Bennett and Shuker, 2017). If an applicant is accepted for a place, they will then participate in a 3–6 month assessment period at Grendon. The assessment takes place in a dedicated wing, which operates as its own therapeutic community for new residents at Grendon (Bennett & Shuker, 2017). The length of the assessment period may depend on how well a resident is settling into the community, whether they have violated any rules, and clinical judgement as to whether the resident is suitable for a therapeutic community. To assess for suitability for Grendon, treatment readiness, motivation, willingness to engage, and adherence to the values of the therapeutic community are evaluated. HMP Grendon is made up of five wings: an assessment wing, a wing for people with sexual offences, a wing for people with mild – moderate learning disabilities (TC+), and three further ‘mainstream’ wings. Each wing can accommodate up to 40 men, apart from TC+ which has a reduced capacity and can hold up to 20 people. Each wing operates as a standalone DTC, but each use the same principles, and share the core components: ‘an

informal atmosphere, regular meetings, resident participation in the running of the community, and residents as auxiliary therapists' (Miller et al., 2006). As described by Bridges (2017), Grendon operates to a weekly timetable. On Monday and Friday mornings, staff and residents hold community meetings where they can discuss ongoing issues, resolve conflicts, and vote on outstanding matters. On Tuesday, Wednesday and Thursday mornings, residents attend small group therapy sessions. Each therapy group is run by clinicians and prison officers and offers residents the opportunity to discuss current conflicts and feelings, work to understand their past, discuss the roots of their offending, and how they might aim to change to prevent offending in future (Brookes, 2010). As well as small group therapy, residents have jobs to support the running of the prison and can take part in psychodrama and art therapy (Jefferies, 2010; Wylie, 2010). The Grendon regime is psychologically intense and requires residents to undertake in-depth therapy about their offending behaviour, upbringing, traumatic events they have experienced, and their current behaviour in prison (Bridges, 2017; Brookes, 2010; Morris, 2004). Qualitative studies with residents about what they value at Grendon, cite the importance of trust, safety, building relationships with staff, and equipping residents with the tools to address and solve problems in a constructive and non-violent way (Dolan, 2017).

Outcome studies have found that residents who are released from Grendon have significantly lower reconviction rates than waiting list controls, and matched controls from mainstream prisons (Marshall, 1997; Taylor, 2000). Research indicates length of stay is important for likelihood of a successful outcome at Grendon, with those who stay at Grendon for 18 months or more, significantly less likely to reoffend than those who stay for shorter durations (Taylor, 2000). As well as reductions in recidivism, research also indicates that post intervention

scores at Grendon are associated with significant improvements in psychological wellbeing, and offence-related risk factors for recidivism (Newton, 1998; Shuker & Newton, 2008). Decreases were found in psychoticism, neuroticism, impulsiveness, and both internally and externally directed hostility (Shuker & Newton, 2008). Additionally, improvements were found in extraversion, self-esteem, and internal locus of control (Newton, 1998). Furthermore, a significant relationship was found between changes in mental health and offence relates risk, and parole board assessments. Shuker and Newton (2008) found that residents who were granted parole after attending HMP Grendon also showed improvements in impulsiveness, neuroticism and self-esteem, compared with those who were not granted parole. This research indicates the substantial changes that can occur after attending HMP Grendon, and adds weight to the evidence supporting the considerable role that Grendon plays in our justice system (Bennett, 2007).

However, Grendon does not work well for everyone. Residents need to be reasonably psychologically minded, motivated to engage in challenging treatment, and able to cope with the rules, regulations of confrontations of this therapeutic environment (Campbell & Attwell, 2018). Dropout rates are significant, and as well as voluntarily leaving, residents can be voted out, known as being ‘deselected’ from therapy for breaking the rules of the DTC (Shuker & Newton, 2008). It is common for over half of admissions to therapeutic communities to not complete the required length of stay for a successful ‘dose’ of treatment which can lead to a large amount of wasted resources (Jones, 1997). This consequently impacts on the facility’s ability to reduce reoffending. In addition, the inclusion of residents who may not end up being suitable for a DTC could significantly impact the culture of the therapeutic community and have a detrimental effect on other resident’s experience. Unfortunately, although attrition

rates are high, little is known about who is likely to succeed at Grendon, and who is not. Shine (2001) conducted research to investigate similarities in characteristics between residents at Grendon, and their length of stay. Shine found two typologies that negatively correlated with length of stay. The first was men who were younger, tough minded, neurotic, and directed hostility externally towards others. The second, was those who were older, emotionally stable, with a frequent history of reoffending. In contrast, they found two typologies who positively correlated with length of stay. The first was those who were neurotic, introverted, and directed hostility internally towards themselves. The second, intelligent, emotionally stable and truthful. This research uses psychometric data to provide us with some indication as to who may do well within therapeutic communities, and who may not. However, this research uses length of stay within a therapeutic community as an outcome which is a proxy measure of success and does not directly tell us whether residents went on to have a successful or non-successful outcome. Therefore, these results must be interpreted with caution, and further research is required to better understand who is likely to have a successful outcome at Grendon, and who is not.

When considering if an individual is suitable for treatment such as Grendon, we must also consider their readiness to engage in treatment. The term ‘treatment readiness’ is frequently used in research to describe whether an individual is likely to successfully engage and benefit from treatment (Ward et al., 2004). There has been considerable research into the role of personality and treatment readiness in forensic settings (Casey et al., 2007; Gaab et al., 2020; Fuller et al., 2019; Howells & Day, 2007; McMurran & Ward, 2010; Ward et al., 2004). One of the most comprehensive and widely used theoretical models of treatment readiness is the Multifactor Offender Readiness Model (MORM; Ward et al., 2004). The MORM proposes

that treatment readiness is a function of both internal and external factors. Internal factors include cognitive, affective, volitional, behavioural and identity factors. The external factors relate to the context and circumstances, such as the location, opportunities, resources, and culture. Ward et al. (2004) proposes that the likelihood of an individual engaging and benefitting from treatment depends upon a combination of both these internal and external factors. Additionally, research to support this model has found that internal factors such as offence-related guilt (Fuller et al., 2019), negative self-evaluation (Alemohammad et al., 2017), interpersonal style (Daffern et al., 2008), and impulsivity (McMurran & Ward, 2010) may play a key role in treatment readiness, and successful outcome of treatment. When considering who is likely to have a successful and non-successful outcome at HMP Grendon, we may need to consider not only the psychological characteristics of the individual, such as personality and mood factors, but also their treatment readiness.

What do we know about research from other DTCs?

As previously mentioned, in the UK Category B male prison estate in England and Wales there are currently three DTCs: HMP Grendon, HMP Gartree and HMP Dovegate. There is also one DTC in the female prison estate, HMP Send. Additional to the research at HMP Grendon, further evidence from research within UK therapeutic communities has found evidence of differing psychometric properties between residents who complete therapy, and those who leave prematurely. Pointon and Roberts (2023) found that when using The Blame Attribution Inventory within a UK prison therapeutic community sample, higher levels of external blame can predict attrition during therapy. Those with higher levels of external attribution are significantly more likely to leave therapy prematurely, including both during

the assessment phase and during core therapy. Additionally, Miller and Brown (2004) studied psychometric properties of residents at HMP Dovegate, and found significant differences between those who are deemed unsuitable for therapy, those who elected to leave during the therapeutic process, and current residents of the DTC. Their research found that leavers judged as unsuitable for therapy were characterised by high scores of Psychopathy, Neuroticism, Criminality, Addiction, Venturesomeness, and Avoidant and Schizoid Personality Disorders. Whereas residents who had elected to leave the TC were more likely to have a sexual index offence, significant scores on the Multiphasic Sex Inventory, and Paranoid and Depressive Personality Disorders. Research regarding DTCs within the UK Forensic context can help inform selection and assessment practices, and should be used to further our knowledge about suitability to attend DTCs. However, the majority of research in other UK DTCs focuses on differences between leavers and current residents of a DTC. Further research is required to investigate differences between residents who are able to achieve a successful outcome after attending a DTC, and those who are not able to achieve a successful outcome.

Aims

This study will examine the relationship between outcome at HMP Grendon, and a variety of psychometric measures which have been demonstrated to have empirical validity as either risk factors for reoffending, indices of treatment readiness or psychological wellbeing. A 'successful' outcome at HMP Grendon will be operationalised as those who have fully met their therapeutic goals, and obtained a progressive move on from Grendon (i.e. moving to a lower security prison or successfully obtaining parole and being released from prison). A

‘non-successful’ outcome will be defined as those who have not met therapeutic goals, and do not have a progressive move on from HMP Grendon. This study aims to identify common profiles that are predictive of successful and non-successful outcome at Grendon. This may also have implications in highlighting treatment needs of individuals who are currently not achieving successful outcomes at Grendon, and those who are at risk of early treatment termination. A fundamental aim of therapeutic communities like Grendon is to reduce recidivism, and aligning individual needs with treatment can increase the likelihood that risk of recidivism is effectively managed (Ward et al., 2004). As Grendon has been demonstrated to reduce reconviction rates (Marshall, 1997; Taylor, 2000), research that aims towards increasing the number of successful outcomes could have vital implications for reducing reoffending rates within the criminal justice system.

Methods

Participants

All participants included in this study were residents at HMP Grendon. All participants were male, and at the time of entering HMP Grendon they were aged between 20 – 78, with a mean age of 35 years (SD = 9.56). All participants included in this study entered HMP Grendon between the years 1988 – 2023. Upon their arrival to HMP Grendon, all residents consented to their data being used anonymously in research (see Appendix 4). Full ethical approval was granted by His Majesty’s Prison and Probation Service and University of Birmingham to conduct this study (see Appendix 1 and 2).

Sampling procedures

A database of anonymised data was created containing demographics and psychometric data from 3060 residents at HMP Grendon. The database also contains information about whether a resident achieved their therapeutic goals, and their onwards move from Grendon.
Whilst attending HMP Grendon, all residents collaboratively set therapeutic goals with clinicians to focus their therapeutic work. Therapeutic goals are created based upon the aim of reducing risk of offending and improving psychological wellbeing. Upon departure of HMP Grendon, the extent to which goals are achieved is determined by a clinician, and is recorded on the database, either as 'not met', 'partially met' or 'fully met'.

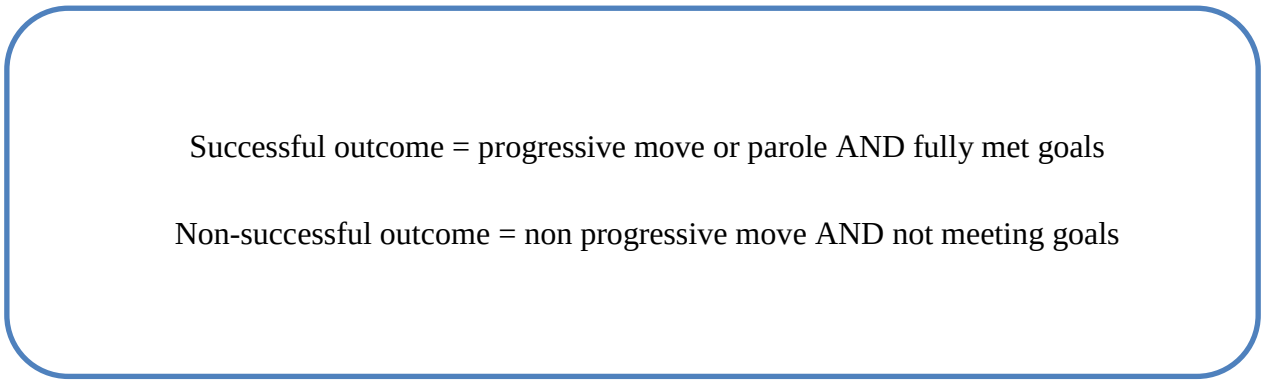
Upon their departure from HMP Grendon, the residents' onwards move is also recorded as a non-progressive move (which is defined by moving on to a prison of the same security rating, a higher security rating, or a psychiatric facility), a progressive move (which is defined as moving on to a lower security prison), or successfully obtaining parole and being released from prison.

Participants were excluded from the study if they were still attending HMP Grendon, or if data was missing for either their onwards move, or for their completion of therapy goals. This reduced the sample size to 2060. Participants were included in the study if they were deemed to have a 'successful' or 'non-successful' outcome after attending HMP Grendon. For the purpose of this study, residents were recorded as having a 'successful' outcome if they successfully met all therapeutic goals whilst at Grendon, AND obtained a progressive move, or parole. A an 'non-successful' outcome was recorded if they did not meet all their therapeutic goals whilst at Grendon, AND if they had a non-progressive move on from

Grendon. This high bar for success was defined as such so that success can more confidently be attributed to the therapeutic process. When looking at only progressive moves from Grendon there is a higher proportion of 'successes' as 29% of individuals gain a progressive onwards move, rather than 19% who meet all therapeutic goals and have a progressive move. However, in this study we define success as both meeting therapeutic goals and progressive moves so we can be more confident that progressive moves have been due to reductions in risk or clinical changes due to the therapeutic process, rather than those who gain a progressive move for other reasons, such as their sentence coming to an end. For ease of reference, the definitions of successful and non-successful outcomes used in this study are summarised in Figure 2.1.

Figure 2.1

Definitions of successful and non-successful outcomes



If participants did not fit into these categories (e.g. those with partially met goals), they were not included in the study as their outcome is an ambiguous mixture of both positive and negative outcomes. Out of the total of 2060 residents, 1244 were deemed to have a non-

successful outcome, and 390 were deemed to have a successful outcome (as highlighted in Table 2.1).

Table 2.1

Outcomes of attending HMP Grendon

		Move type			
		Non-Progressive Move	**Progressive Move**	**Parole**	**Total**
Meeting therapeutic goals	Not met	1244	9	3	1256
	Partially met	215	159	35	409
	Fully met	5	354	36	395
Total		1464	522	74	2060

Note This table includes information about whether residents fully met, partially met, or did not meet their therapeutic goals. It also presents whether participants had a non-progressive move on from Grendon (defined as not moving on to a lower security prison) a progressive move (moving to a lower security prison) or achieved parole. The highlighted sections were those categorised for the purpose of this research as non-successful' in yellow, and successful' in blue.

The participants selected for analysis were aged between 21 – 77, with a mean age of 36 years (SD = 10). The participants entered Grendon between the years of 1988 – 2019. The ethnicities of participants included in this study are listed in Table 2.2. Please note, due to missing data, ethnicities are provided for 1145 out of a total of 2060 participants.

Table 2.2

Ethnicities of participants selected to take part in the study

		Number of participants	Percentage
Ethnicity	White British	865	75.5%
	White Irish	21	1.8%
	White Other	27	2.4%
	Asian Indian	6	0.5%
	Asian Pakistani	24	2.1%
	Asian Bangladeshi	3	0.3%
	Asian Other	12	1.0%
	Black Caribbean	90	7.9%
	Black African	21	1.8%
	Black Other	22	1.9%
	Mixed White and Black Caribbean	26	2.3%
	Mixed White and Black African	7	0.6%
	Mixed White and Asian	8	0.7%
	Mixed other	11	1.0%
	Other	2	0.2%
Total		1145	100.0%

Of the participants included in this study, 202 were completing discretionary life sentences, 448 were completing mandatory life sentences, 297 were completing Imprisonment for Public Protection (IPP) sentences, and 472 were completing determinate sentences. The most frequent index offences were murder (236 participants), robbery (121 participants), rape (106 participants), grievous bodily harm (59 participants), attempted murder (46 participants), and manslaughter (18 participants). Other index offences include armed robbery, burglary, attempted rape, sexual assault, kidnap, arson, conspiracy to murder.

Measures

A battery of psychometric test was completed with participants upon arrival to HMP Grendon. These psychometric measures are described below:

Personality Disorder Questionnaire

The Personality Disorder Questionnaire (PDQ4; Hyler, 1994) is a 99 item self-report questionnaire and is used to screen for the presence of personality disorders as found in the

Diagnostic and Statistical Manual Fourth edition (DSM IV; American Psychiatric Association, 2000). The questionnaire provides an indication as to whether the individual meets the clinical threshold for a personality disorder, and the subscales are Schizoid, Schizotypal, Depressive, Histrionic, Narcissistic, Borderline, Antisocial and Conduct Disorder, Avoidant, Dependent, and Obsessive Compulsive.

Personality Assessment Inventory

The Personality Assessment Inventory (PAI; Morey, 1991) is a 344 item self-report questionnaire that assesses psychopathological syndromes and personality traits. It forms 22 nonoverlapping scales: 4 validity scales (Inconsistency, Infrequency, Negative Impression, Positive Impression), 11 clinical scales (Somatic Complaints, Anxiety, Anxiety Related Disorders, Depression, Mania, Paranoia, Schizophrenia, Borderline Features, Antisocial Features, Alcohol Problems, Drug Problems), 5 treatment consideration scales (Aggression, Suicidal Ideation, Stress, Non-support, Treatment Rejection) and 2 interpersonal scales (Dominance and Warmth). The PAI has been found to have good reliability and validity (Busse et al., 2014), with a mean Cronbach's alpha of the full subscales reported to be .86 (Morey, 1991).

Hostility and Direction of Hostility

Hostility and Direction of Hostility Questionnaire (HDHQ; Caine et al., 1967) is a 51 item self-report measure, and contains 5 subscales Urge to Act Out Hostility (AH), Criticism of Others (CO), Projected Delusional or Paranoid Hostility (PH), Self-Criticism (SC) and Guilt (G). The first 3 subscales are summed to form an Extrapunitive (EH) score, which indicates hostility is being directed externally towards others. The latter two subscales are added to

yield an Intropunitive (IH) score, which indicates hostility is being directed internally, towards the self. A total Direction of Hostility (DH) score is also calculated by taking the sum of the 3 Extrapunitive scales (AH + CO + PH) and subtracting it from the sum of twice the SC score and the G score: Direction of Hostility (DH) = (2SC + G) - EH). Positive scores thus indicate the extent of intropunitive hostility, with higher scores indicating a higher level of intropunitive hostility. Negative scores indicate extrapunitive hostility, with lower scores indicating higher levels of extrapunitive hostility. The measure has been shown to have adequate reliability and validity (Arrindell et al., 1984).

Psychological Inventory of Criminal Thinking Styles

The Psychological Inventory of Criminal Thinking Styles (PICTS; Walters, 1995) is an 80 item self-report measure designed to assess for cognitive styles of criminal thinking. The PICTS contains eight subscales, providing measures of Mollification (assigning the cause of behaviours to external factors), Entitlement (an attitude of ownership and incorrectly assigning wants as needs), Power orientation (engaging in behaviour designed to control and manipulate others), Sentimentality (compensating for past negative behaviour by engaging in good deeds), Super Optimism (feeling invincible, believing that they have the ability to not get caught, or avoid consequences of committing crime), Cognitive Indolence (a tendency to have an uncritical view of their plans and ideas, a focus on short term gains, and short-cut problem solving), Discontinuity (disruption of thought and inability to follow things through) and Cutting off (the rapid elimination of common psychological deterrents to crime). The measure also contains two validity scales: Confusion (where someone appears to have difficulty understanding the items) and Defensiveness (lack of acknowledgement or insight

regarding difficulties that arise from a criminal lifestyle). The PICTS is reported to have good reliability and validity (Walters, 1995).

Gudjonsson Blame Attribution Inventory - Revised

The Gudjonsson Blame Attribution Inventory - Revised (GBAI-R; Gudjonsson & Singh, 1989) is a 42 item self-reported questionnaire, which measures an individual's blame attribution. Blame attribution refers to the concept of attempting to construct causal explanations for behaviours that they, or others, might display. The GBAI-R contains three subscales, Guilt attribution (the extent to which an individual feels guilt and remorse for their crimes), External attribution (the extent to which an individual blame the victims or society for their crime), and Mental attribution (the extent to which an individual states they had no mental control). Cronbach's alpha has been reported to be between .67–.89. (Cima et al., 2007; Fox et al., 2003; Weizmann-Henelius et al., 2002).

Corrections Victoria Treatment Readiness Questionnaire

The Corrections Victoria Treatment Readiness Questionnaire (CVTRQ; Casey et al., 2007) is used to measure an individuals' readiness to engage in treatment. The CVTRQ is a 20-item self-report questionnaire and includes four subscales: Attitudes and Motivation (6 items measuring attitudes and beliefs about programs, and motivation to change), Emotional Reactions (6 items measuring emotional responses to the individual's offending behaviour), Offending Beliefs (4 items measuring the individual's beliefs about personal responsibility for offending) and Efficacy (4 items measuring the individual's perceived ability to participate in programs). Total scores on the CVTRQ are also calculated and can range from 40 to 200, with higher scores indicating a higher degree of treatment readiness. Casey et al. (2007) reports acceptable levels of reliability and validity.

Persons Relating to Others Questionnaire

The Persons Relating to Others Questionnaire (PROQ3; Birtchnell et al., 2013) is used to assess an individual's relational style. The PROQ3 is a 48-item self-report questionnaire, comprising of 8 subscales: Upper Distant (sadistic, intimidating and tyrannising), Upper Neutral (pompous, boastful, dominating and insulting), Upper Close (intrusive, restrictive, possessive), Neutral Distant (suspicious, uncommunicative, self-reliant), Neutral Close (fear of separation, fear of being alone), Lower Distant (acquiescent, subservient, withdrawn), Lower Neutral (helpless, shunning, responsibility, self-denigrating) and Lower Close (fear of rejection and disapproval). It has been shown to have acceptable Cronbach's alpha coefficients across a variety of samples and has demonstrated convergent and discriminant validity (Birtchnell et al., 2013).

Culture-Free Self-Esteem Inventory

Culture-Free Self-Esteem Inventory (CFSEI-2; Battle, 1992) is a 40-item self-report inventory, consisting of four subscales: General Self-Esteem, Social Self-Esteem, Personal Self-Esteem, and a Lie subtest. It has been demonstrated to have a test re-test reliability of 0.81, and good validity (Battle, 1992).

Analysis Strategy

The first stage of this analysis will focus on the identification of scales and subscales that are systematically associated with successful or non-successful treatment outcomes. Due to difficulties with missing data and co-linearity between predictors, this stage will be completed as a series of independent logistic regression analyses. A forward stepwise selection method was used to select the scales within a measure that are most predictive of treatment outcome.

After a set of potential predictor variables have been identified, then a classification and regression tree (CRT) analysis will be conducted to formulate a series of decision rules using these putative predictor variables that aim to identify those likely to present with successful and non-successful treatment outcome from data obtained prior to engaging with the Grendon treatment programme.

Results

The Impact of Meeting the Threshold for Personality Disorders

The PDQ4 (Hyler, 1994) is used as a screening tool for the presence DSM IV Axis II personality disorders. Frequency counts for the PDQ subscales are described for successful and non-successful outcomes in Table 2.3. Values in the same row that do not share the same subscript are significantly different at $p < 0.05$ in a two-sided test of proportions. Tests are adjusted for all pairwise comparisons using the Benjamini-Hochberg correction.

Table 2.3

Frequency statistics for the Personality Diagnostic Questionnaire

PDQ Subscales	Threshold	Non-successful	Successful
Antisocial and Conduct Disorder	Below threshold	83_a	46_a
	Above Threshold	133_a	62_a
Schizoid Personality Disorder	Below threshold	189_a	98_a
	Above Threshold	40_a	11_a
Schizotypal Personality Disorder	Below threshold	175_a	93_a
	Above Threshold	54_a	16_a
Paranoid Personality Disorder	Below threshold	108_a	57_a
	Above Threshold	121_a	52_a
Avoidant Personality Disorder	Below threshold	124_a	62_a
	Above Threshold	105_a	47_a
Dependent Personality Disorder	Below threshold	212_a	103_a
	Above Threshold	17_a	6_a
Obsessive Compulsive Personality Disorder	Below threshold	149_a	73_a
	Above Threshold	80_a	36_a
Histrionic Personality Disorder	Below threshold	208_a	106_b

PDQ Subscales	Threshold	Non-successful	Successful
	Above Threshold	21_a	3_b
Narcissistic Personality Disorder	Below threshold	194_a	93_a
	Above Threshold	35_a	16_a
Borderline Personality Disorder	Below threshold	116_a	59_a
	Above Threshold	102_a	49_a

These potential predictor variables were entered into a forward conditional stepwise logistic regression. The stepwise procedure resulted in a significant logistic regression equation (X^2 = 5.322, p =0.021) containing one predictor variable (see Table 2.4).

Table 2.4

Personality Diagnostic Questionnaire predictors retained in the forward conditional stepwise logistic regression

							95% C.I. for Exp(B)	
PDQ Subscales	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Histrionic Personality Disorder	-1.273	0.631	4.072	1	0.044	0.28	0.081	0.964
Constant	-0.624	0.121	26.636	1	<.001	0.536		

The Exp(B) value describes the change in the odds of successful outcome at HMP Grendon given the presence of the personality diagnosis. Exp(B) values greater than one being associated with successful outcome and values less than one being associated with non-successful outcome (i.e., percentage change in odds = ((1– Exp(B)) * 100). Therefore, assuming that all other variables are held constant, the odds of a successful outcome at Grendon decreased by 72% given the presence of a histrionic personality disorder diagnosis.

The Impact of the Personality Factors

Personality factors were measured on the PAI (Morey, 1991). The PAI should be distinguished from the PDQ4 in that it measures different personality traits, rather than those that meet a threshold for a specific personality disorder.

Descriptive statistics for the subscales of the PAI are described in Table 2.5. Values in the same row that do not share the same subscript are significantly different at p< .05 in the two-sided test of equality for column means.

Table 2.5

Descriptive statistics for the subscales of the Personality Assessment Inventory

	Non-successful outcome			Successful outcome		
PAI Subscales	**Mean**	**Standard Deviation**	**Count**	**Mean**	**Standard Deviation**	**Count**
Inconsistency scale (Validity)	56a	10	1244	55b	9	390
Infrequency scale (Validity)	60a	10	1244	58a	9	390
Negative Impression scale (Validity)	62a	19	1244	58b	15	390
Positive Impression scale (Validity)	45a	12	1244	51b	33	390
Somatic Complaints scale (Clinical)	54a	11	1244	50b	8	390
Anxiety scale (Clinical)	59a	13	1244	56a	12	390
Anxiety Related Disorders scale (Clinical)	61a	14	1244	60a	13	390
Depression scale (Clinical)	63a	14	1244	60b	13	390
Mania scale (Clinical)	49a	12	1244	45b	10	390
Paranoia scale (Clinical)	64a	14	1244	59b	13	390
Schizophrenia scale (Clinical)	60a	16	1244	55b	13	390
Borderline Features scale (Clinical)	69a	38	1244	63b	14	390
Antisocial Features scale (Clinical)	70a	13	1244	64b	12	390
Alcohol Problems scale (Clinical)	62a	18	1244	61a	18	390
Drug Problems scale (Clinical)	69a	20	1244	66b	19	390
Aggression scale (Treatment)	63a	15	1244	58b	14	390
Suicidal Ideation scale (Treatment)	62a	19	1244	58b	15	390
Stress scale (Treatment)	59a	12	1244	58a	11	390
Non-support scale (Treatment)	61a	14	1244	59a	13	390
Treatment Rejection scale (Treatment)	35a	9	1244	34a	9	390
Dominance scale (Interpersonal)	48a	11	1244	45b	10	390
Warmth scale (Interpersonal)	42a	11	1244	44b	10	390

Note Values in the same row that do not share the same subscript are significantly different at p<0.05 in a two-sided test of mean difference. Tests are adjusted for all pairwise comparisons using the Benjamini-Hochberg correction.

These potential predictor variables were entered into a forward conditional stepwise logistic regression. The stepwise procedure resulted in a significant logistic regression equation (X^2 = 62.423, p <.001) containing six predictor variables. The predictors retained in the final equation are presented in Table 2.6.

Table 2.6

Personality Assessment Inventory predictors retained in the forward conditional stepwise logistic regression

							95% C.I. for Exp(B)	
PAI Subscales	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Somatic Complaints scale (Clinical)	-0.036	0.011	9.876	1	0.002	0.965	0.943	0.987
Anxiety Related Disorders scale (Clinical)	0.022	0.009	5.339	1	0.021	1.022	1.003	1.041
Antisocial Features scale (Clinical)	-0.025	0.009	7.733	1	0.005	0.975	0.958	0.993
Suicidal Ideation scale (Treatment)	-0.017	0.007	6.342	1	0.012	0.983	0.97	0.996
Treatment Rejection scale (Treatment)	-0.027	0.012	5.068	1	0.024	0.973	0.951	0.996
Dominance scale (Interpersonal)	-0.021	0.009	4.947	1	0.026	0.979	0.962	0.998
Constant	3.297	1.289	6.538	1	0.011	27.024		

The Exp(B) value describes the change in the odds of a successful outcome at HMP Grendon, with Exp(B) values greater than 1 being associated with a successful outcome, and values less than 1 being associated with a non-successful outcome (i.e., percent change in odds = (1-Exp(B))*100). Therefore, assuming that all other variables are held constant, the odds of a successful outcome at Grendon increased by 2.2% for every one unit increase in Anxiety Related Disorders scale.

The odds of a non-successful outcome at Grendon increased by 3.6% for a one-unit increase in the Somatic Complaints scale, 2.5% for the Antisocial Features scale, 1.7 % for the Suicidal Ideation scale, 2.7% for the Treatment Rejection scale, and 2.1% for the Dominance scale.

The Impact of Hostility

Hostility was measured on HDHQ (Caine et al., 1967). Mean and standard deviations for the HDHQ subscales are described for successful and non-successful outcomes in Table 2.7.

Table 2.7

Descriptive statistics for the subscales of the Hostility and Direction of Hostility Questionnaire

	Non-successful			Successful		
HDHQ Subscales	**Mean**	**Standard Deviation**	**Count**	**Mean**	**Standard Deviation**	**Count**
Self-Criticism	6.1a	2.7	1244	6.1a	2.7	390
Guilt	3.9a	1.9	1244	3.9a	1.8	390
Urge to Act Out Hostility	5.9a	2.6	1244	5.3b	2.4	390
Projected Delusional or Paranoid Hostility	2.9a	2.0	1244	2.3b	1.6	390
Criticism of Others	5.7a	2.9	1244	4.9b	2.8	390
Total HDHQ Score	24.5a	8.9	1244	22.5b	8.6	390
Direction of Hostility	1.8a	7.3	1244	3.5b	6.4	390
Intropunitive Hostility	10.1a	4.2	1244	9.9a	4.1	390
Extrapunitive Hostility	14.4a	6.4	1244	12.5b	5.8	390

Note Values in the same row that do not share the same subscript are significantly different at p<0.05 in a two-sided test of mean difference. Tests are adjusted for all pairwise comparisons using the Benjamini-Hochberg correction.

These potential predictor variables were entered into a forward conditional stepwise logistic regression. The stepwise procedure resulted in a significant logistic regression equation (X^2 = 25.759, $p < 0.001$) containing one predictor variable (see Table 2.8).

Table 2.8

Hostility and Direction of Hostility predictors retained in the forward conditional stepwise logistic regression

							95% C.I. for Exp(B)	
HDHQ Subscales	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Extrapunitive Hostility	-0.050	0.010	24.855	1	0.000	0.951	0.933	0.970
Constant	-0.332	0 143	5 366	1	0.021	0.718		

Assuming that all other variables are held constant, the odds of a successful outcome at Grendon decreased by 5% for every unit increase on the extra unit increase in the Extrapunitive hostility subscale of the HDHQ.

The Impact of Criminal Thinking Style

Criminal Thinking Style was measure using the PICTS (Walters, 1995). Mean and standard deviations for the PICTS subscales are described for successful and non-successful outcomes at HMP Grendon in Table 2.9.

Table 2.9

Descriptive statistics for the Psychological Inventory of Criminal Thinking Styles

	Non-successful			Successful		
PICTS Subscales	**Mean**	**Standard Deviation**	**Count**	**Mean**	**Standard Deviation**	**Count**
Confusion	52.66_a	12.42	1244	48.70_b	10.69	390
Defensiveness	44.22_a	9.11	1244	42.41_b	7.08	390
Mollification	49.32_a	10.80	1244	46.59_b	9.35	390
Cut-Off	54.34_a	10.67	1244	52.25_b	10.16	390
Entitlement	51.84_a	11.12	1244	48.76_b	9.37	390
Power orientation	51.58_a	11.51	1244	48.64_b	10.89	390
Sentimentality	48.69_a	10.85	1244	46.32_b	8.76	390
Super-optimism	52.80_a	11.59	1244	50.55_b	10.10	390
Cognitive indolence	51.25_a	11.29	1244	49.63_b	10.83	390
Discontinuity	52.06_a	10.70	1244	50.23_b	10.68	390

Note Values in the same row that do not share the same subscript are significantly different at p<0.05 in a two-sided test of mean difference. Tests are adjusted for all pairwise comparisons using the Benjamini-Hochberg correction.

These potential predictor variables were entered into a forward conditional stepwise logistic regression. The stepwise procedure resulted in a significant logistic regression equation (X^2 = 42.276, $p < 0.001$) containing three predictor variables (see Table 2.10).

Table 2.10

Psychological Inventory of Criminal Thinking Styles predictors retained in the forward conditional stepwise logistic regression

							95% C.I. for Exp(B)	
PICTS Subscales	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Confusion	-0.028	0.007	15.932	1	0.000	0.972	0.959	0.986
Defensiveness	-0.036	0.010	13.860	1	0.000	0.965	0.947	0.983
Entitlement	-0.020	0.008	6.186	1	0.013	0.980	0.965	0.996

Constant	2.812	0.663	18.013	1	0.000	16.648

Assuming that all other variables are held constant, the odds of a successful outcome at Grendon decreased by 2.8% for every unit increase in the Confusion subscale, 3.6% for the Defensiveness subscale and 2% for every unit increase in the Entitlement subscale. Both Confusion and Defensiveness are validity scales of the PICTS.

The Impact of Blame Attribution

The GBAI-R (Gudjonsson & Singh, 1989) was used to measure an individual's attribution of blame. Mean and standard deviations for the GBAI-R subscales are described for successful and non-successful outcomes at HMP Grendon in Table 2.11.

Table 2.11

Descriptive statistics for the Gudjonsson Blame Attribution Inventory – Revised

	Non-successful			Successful		
BAI Subscales	**Mean**	**Standard Deviation**	**Count**	**Mean**	**Standard Deviation**	**Count**
Mental attribution	4.9_a	2.5	1244	5.1_a	2.2	390
External attribution	1.8_a	2.4	1244	1.6_a	1.9	390
Guilt attribution	12.2_a	4.0	1244	12.9_b	3.5	390

Note Values in the same row that do not share the same subscript are significantly different at p<0.05 in a two-sided test of mean difference. Tests are adjusted for all pairwise comparisons using the Benjamini-Hochberg correction.

These potential predictor variables were entered into a forward conditional stepwise logistic regression. The stepwise procedure resulted in a significant logistic regression equation (X^2 = 7.969, p <.005) containing one predictor variable (see Table 2.12).

Table 2.12

Gudjonsson Blame Attribution Inventory predictors retained in the forward conditional stepwise logistic regression

GBAI-R subscales	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for Exp(B)	
							Lower	Upper
Guilt	0.044	0.016	7.718	1	0.005	1.045	1.013	1.079
Constant	-1.602	0 212	57.180	1	0.000	0.201		

Assuming that all other variables are held constant, the odds of a successful outcome at Grendon increased by 4.4% for every unit increase in the Guilt subscale.

The Impact of Treatment Readiness

Treatment Readiness was measured using the CVTRQ (Casey et al., 2007). Mean and standard deviations for the CVTRQ subscales are described for successful and non-successful outcomes at HMP Grendon in Table 2.13. Total scores on the CVTRQ are also calculated and can range from 40 to 200, with higher scores indicating a higher degree of treatment readiness.

Table 2.13

Descriptive statistics for the Corrections Victoria Treatment Readiness Questionnaire

CVTRQ Subscales	Non-successful			Successful		
	Mean	Standard Deviation	Count	Mean	Standard Deviation	Count
Attitudes and Motivation	26_a	4	1244	28_b	3	390
Emotional Reactions	24_a	5	1244	25_b	4	390
Offender Beliefs	17_a	4	1244	18_b	3	390
Efficacy	14_a	3	1244	15_b	3	390
Total CVTRQ Score	81_a	12	1244	85_b	8	390

Note Values in the same row that do not share the same subscript are significantly different at p<0.05 in a two-sided test of mean difference. Tests are adjusted for all pairwise comparisons using the Benjamini-Hochberg correction.

These potential predictor variables were entered into a forward conditional stepwise logistic regression. The stepwise procedure resulted in a significant logistic regression equation (X^2 = 16.502, p < .001) containing one predictor variable (see Table 2.14).

Table 2.14

Corrections Victoria Treatment Readiness predictors retained in the forward conditional stepwise logistic regression

						95% C.I. for Exp(B)		
Predictors	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Total CVTRQ Score	0.042	0.011	13.743	1	0.000	1.043	1.020	1.067
Constant	-4.788	0 970	24.350	1	0.000	0.008		

Assuming that all other variables are held constant, the odds of a successful outcome at Grendon increased by 4.2% for every unit increase in the Total CVTRQ Score.

The Impact of Relational Style

The PROQ3 (Birtchnell et al., 2013) was used to assess the impact of an individual's relational style on the likelihood of a successful outcome at HMP Grendon. Mean and standard deviations for the PROQ3 subscales are described for successful and non-successful outcomes at HMP Grendon in Table 2.15.

Table 2.15

Descriptive statistics for the Persons Relating to Others Questionnaire

	Non-successful			Successful		
PROQ3 Subscales	**Mean**	**Standard Deviation**	**Count**	**Mean**	**Standard Deviation**	**Count**
Upper Neutral score	5a	4	1244	5a	3	390
Upper Close score	4a	4	1244	3a	4	390
Neutral Close score	4a	4	1244	4a	4	390
Lower Close score	6a	4	1244	7a	4	390

Lower Neutral score	6a	4	1244	6a	4	390
Lower Distant score	5a	4	1244	6a	4	390
Neutral Distant score	8a	4	1244	7b	4	390
Upper Distant score	7a	4	1244	6b	4	390
Total PROQ3 Score	45a	20	1244	43a	19	390

Note Values in the same row that do not share the same subscript are significantly different at p<0.05 in a two-sided test of mean difference. Tests are adjusted for all pairwise comparisons using the Benjamini-Hochberg correction.

These potential predictor variables were entered into a forward conditional stepwise logistic regression. The stepwise procedure resulted in a significant logistic regression equation (X^2 = 19.293, p < .001) containing three predictor variables (see Table 2.16).

Table 2.16

Persons Relating to Others Questionnaire predictors retained in the forward conditional stepwise logistic regression

							95% C.I. for Exp(B)	
PROQ3 Subscales	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Upper Close score	-0.052	0.022	5.718	1	0.017	0.950	0.911	0.991
Lower Close score	0.069	0.021	10.414	1	0.001	1.071	1.027	1.117
Neutral Distant score	-0.070	0.019	13.781	1	0.000	0.932	0.898	0.967
Constant	-0.842	0.149	32.112	1	0.000	0.431		

Assuming that all other variables are held constant, the odds of a successful outcome at HMP Grendon increased by 6.9% for every unit increase in the Lower Close score (fear of rejection and disapproval). The odds of a non-successful outcome at HMP Grendon increased by 5.2% for every unit increase in the Upper Close score (intrusive, restrictive, possessive), and 7% for every unit increase in the Neutral Distant score (suspicious, uncommunicative, self-reliant).

The Impact of Self-Esteem

To measure the impact of Self-Esteem, the CFSEI-2 (Battle, 1992) was used. Mean and standard deviations for the CFSEI-2 subscales are described for successful and non-successful outcomes at HMP Grendon in Table 2.17.

Table 2.17

Descriptive statistics for the Persons Relating to Others Questionnaire

	Non-successful			Successful		
CFSEI subscales	**Mean**	**Standard Deviation**	**Count**	**Mean**	**Standard Deviation**	**Count**
Social Self-Esteem	4.8_a	2.1	1244	5.3_b	2 1	390
Personal Self-Esteem	3.6_a	2.6	1244	4.1_b	2.6	390
Lie Subtest	1.6_a	1.5	1244	$1\ 5_a$	1 5	390
General Self-Esteem	8.3_a	4.2	1244	8.7_a	4.0	390

These potential predictor variables were entered into a forward conditional stepwise logistic regression. The stepwise procedure resulted in a significant logistic regression equation (X^2 = 11.496, $p < .001$) containing one predictor variable (see Table 2.18).

Table 2.18

Culture-Free Self-Esteem Inventory predictors retained in the forward conditional stepwise logistic regression

							95% C.I. for Exp(B)	
CFSEI subscales	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Social Self-Esteem	0.103	0.031	11.176	1	0.001	1.109	1.044	1.178
Constant	-1.518	0.171	78.901	1	0.000	0.219		

Assuming that all other variables are held constant, the odds of a successful outcome at HMP Grendon increased by 10.9% for every unit increase in the Social Self-Esteem subscale.

Clinical Decision Tree

16 variables were identified by the previous logistic regressions as predictive of successful or non-successful outcome, as summarised in Table 2.19.

Table 2.19

Summary of significant predictors identified in stepwise logistic regressions

Measure	Subscale
Personality Disorder Questionnaire	Anxiety Related Disorders
Gudjonsson Blame Attribution Inventory - Revised	Guilt
Corrections Victoria Treatment Readiness Questionnaire	Total CVTRQ score
Persons Relating to Others Questionnaire	Lower Close
Culture-Free Self-Esteem Inventory	Social Self-Esteem
Personality Disorder Questionnaire	Histrionic Personality Disorder
Personality Assessment Inventory	Somatic Complaints Antisocial Features Suicidal Ideation Treatment Rejection Dominance
Hostility and Direction of Hostility Questionnaire	Extrapunitive Hostility
Psychological Inventory of Criminal Thinking Styles	Confusion Defensiveness Entitlement
Persons Relating to Others Questionnaire	Upper Close Neutral Distant

Unfortunately, none of the participants had a complete dataset for all 16 variables. However, if the variable relating to Histrionic Personality Disorder (PDQ4) was omitted from the

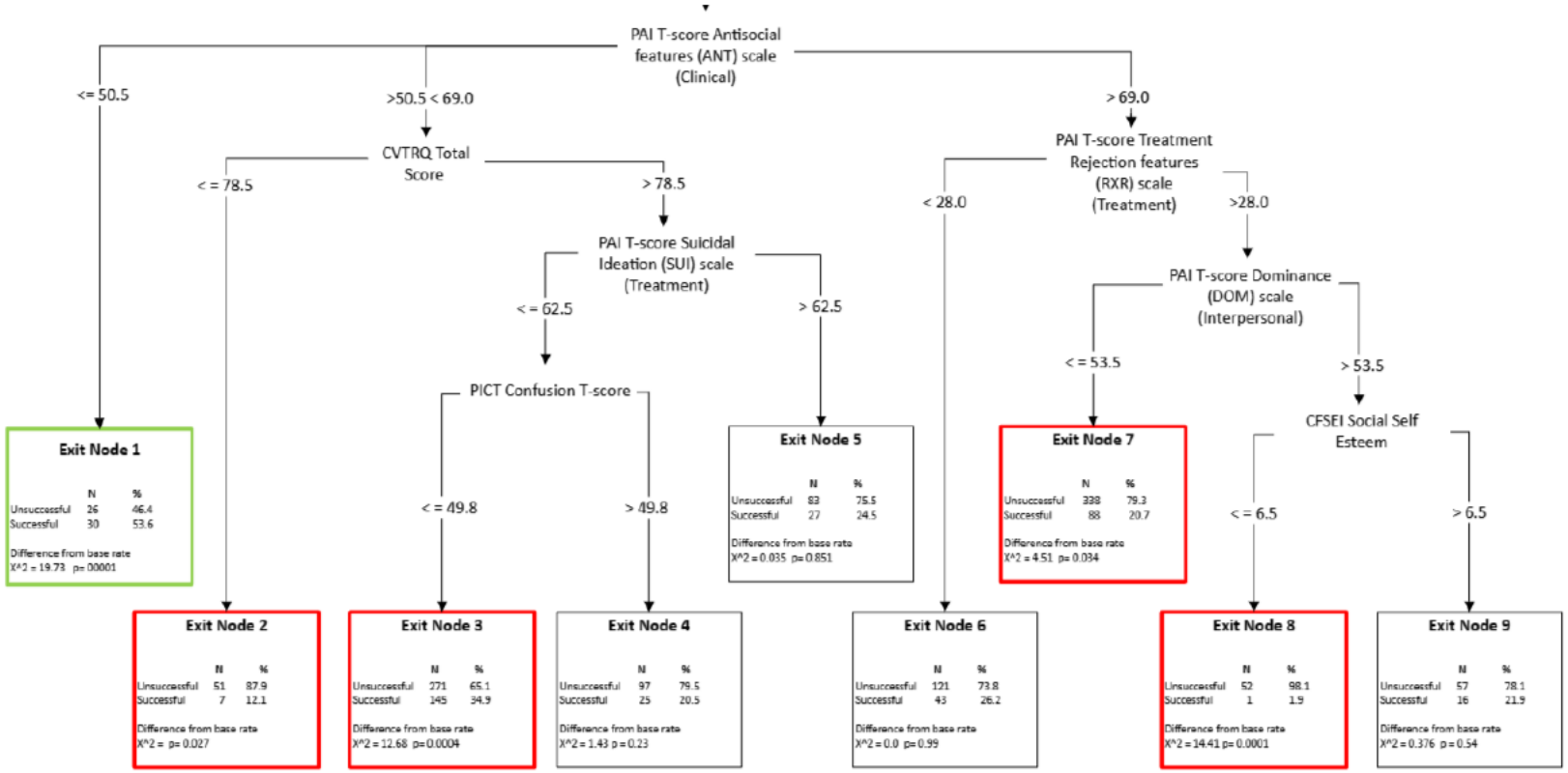
dataset, then 1478 participants had a complete dataset for the remaining 15 variables. Therefore, the clinical decision tree was calculated from the 1478 participants who had valid data on of all the 15 potential classification variables.

A classification and regression tree analysis (CRT) was conducted to determine how the measures identified as potential predictors of outcome could be used to inform clinical decisions regarding who is likely to present successful and non-successful treatment outcomes. Unlike the prior logistic regressions, CRT does not develop a predictive equation. Instead, classification rules are formulated using the potential dependent variables in the form of a decision tree. The classification rules are formulated to maximize correct classification of cases and the decision tree that can be used to make predictions from new observations.

The CRT analysis was conducted with the prior probabilities obtained from the frequency of the successful and non-successful outcomes as observed in the dataset. The "Gini" algorithm was used to generate 'decisions' that maximise the homogeneity of the exit nodes with respect to successful and non-successful outcomes. The CRT decision tree is shown in Figure 2.2, with highlighted exit nodes demonstrating statistically significant different rates of successful and non-successful outcomes relative to base rate estimates.

Figure 2.2

Clinical Decision Tree



Note CRT decision tree to maximize correct classification of cases in terms of successful or non-successful treatment. Highlighted exit nodes show a statistically significant different rate of successful and non-successful outcomes relative to base rate estimates, with red highlights indicating a greater rate of non-successful treatment and green highlights indicating greater rights of successful treatment.

As can be seen in Figure 2.2, there were four exit nodes (nodes 2, 3, 7 and 8) that resulted in rates of non-successful outcome that significantly exceeded base rate estimates. For example, scoring a greater t-score than 69 on the Antisocial subscale of the PAI, greater than 28 on the Treatment Rejection subscale (PAI), greater than 53.5 on the Dominance subscale (PAI), and less than 6.5 on the Social Self Esteem subscale of the CFSEI was associated with 98.1% non-successful treatment rate. Only one exit node (node 1) resulted in the rate of successful outcome that significantly exceeded base rate estimates, with this occurring in participants scoring a t-score less than 50.5 on the Antisocial subscale of the PAI.

The decision rules for successful and non-successful treatment are described in Table 2.20.

Table 2.20

The decision rules for successful and non-successful treatment

Exit node	Predicts	Decision Rule
1	Successful	T-score for PAI Antisocial Features Subscale < 50.5
2	Non-successful	T-score for PAI Antisocial Features Subscale > 50.5 & < 69 AND CVTRQ Total score < = 78.5
3	Non-successful	T-score for PAI Antisocial Features Subscale > 50.5 & < 69 AND CVTRQ Total Score > 78.5 AND T-score for PAI Suicidal Ideation <62.5 AND T-score for PICT Confusion Subscale < 49.8
7	Non-successful	T-score for PAI Antisocial Features Subscale > 69 AND T-score for PAI Treatment Rejection Subscale > 28 AND T-score for PAI Dominance subscale < 53.5
8	Non-successful	T-score for PAI Antisocial Features Subscale > 69 AND T-score for PAI Treatment Rejection Subscale > 28 AND T-score for PAI Dominance Subscale > 53.5 AND CFSEI Self Esteem < 6.5

The classification accuracy of this decision tree is shown in Table 2.21. An overall correct classification rate of 74.4% was observed, with 97.6% of non-successful outcome correctly predicted. However, it should be noted that only 7.9% of successful outcomes were predicted correctly.

Table 2.21

Classification accuracy of the CRT decision tree

	Predicted		
Observed	**Non-successful**	**Successful**	**Percent Correct**
Non-successful	1070	26	97.6%
Successful	352	30	7.9%
Overall Percentage	96.2%	3.8%	74.4%

This suggests that the decision tree has greater power to predict non-successful outcomes than to predict successful outcomes.

Discussion

This study outlines a model that can be used to predict non-successful outcomes at HMP Grendon with an overall accuracy of 97.6% (Table 2.21). When using the clinical regression tree, the most accurate prediction of non-successful outcomes is Exit Node 8 (Table 2.20) which describes individuals who self-report higher levels of Antisocial Features and Treatment Rejection on the PAI, higher Dominance in the PROQ3, and lower Social Self-Esteem on the CFSEI. With this profile, the model can predict a non-successful outcome with a 98.1% accuracy. This is supported by previous research that indicates that antisocial

personality features have been found to be a predictor for treatment non-completion (Holdsworth et al., 2014; Pelissier et al., 2003). Additionally, Watson et al. (2017) found that offenders who rated highly in a dominant interpersonal style, were more likely to report major ruptures in the therapeutic alliance, which may consequently impact outcome in the therapeutic community (Bender, 2005; Samstag et al., 1998). Previous research regarding self-esteem in therapeutic communities found no significant differences between self-esteem in current residents, and those who were deemed unsuitable for treatment, or who were deselected (Miller et al., 2004). However, our study demonstrates that those who have a non-successful outcome at Grendon have significantly lower social self-esteem than those who achieve successful outcomes after attending Grendon. Low self-esteem is known to be correlated with depression, anxiety, social phobia, anorexia, bulimia, body-dysmorphic disorder and alcohol abuse (Zeigler-Hill, 2013) and such co-morbidity may increase complexity of treatment required. Additionally, Schanz (2017) found that inmates with low self-esteem were more likely to display physical passive aggression such as drug abuse or self-destructive behaviour, which would violate the core rules of the therapeutic community and lead to early treatment termination. When considering specifically low social self-esteem, research has shown a significant correlation with low social self-esteem and social anxiety (Valentiner et al., 2011) and this is likely to interfere with residents' ability to function in a social environment such as Grendon. Building therapeutic alliances, friendships, speaking in group therapy sessions, community meetings, and taking part in psychodrama are all likely to be impacted if a resident is experiencing low social self-esteem. Consequently, residents may be seen as simply not 'engaging' with treatment which may cause negative appraisals from staff members.

This study suggests that improving social self-esteem may help improve outcomes at Grendon. When comparing Exit Node 8 and 9 (Table 2.20), all individuals show higher levels in Antisocial Features and Treatment Rejection on the PAI and higher Dominance in the PROQ3. However, more residents gain a successful outcome at Grendon when higher in Social Self-Esteem, rather than those who are lower in Social Self-Esteem. 21.9% of people with this profile achieve a successful outcome at Grendon if Social Self-Esteem is above 6.5 on the CFSEI, compared with just 1.9% of people who achieve a successful outcome at Grendon when their Social Self-Esteem is below 6.5. If identified at assessment, this highlights a target for therapy which may improve the likelihood of gaining a successful outcome at Grendon for residents who self-report high levels of Antisocial Features, Treatment Rejection and Dominance.

The second highest predictor of non-successful outcomes were those with a profile of moderate Antisocial Features (between 50.5-69 on the PAI) and lower levels of Treatment Readiness (below 78.5 on the CVTRQ). This is consistent with extensive research into treatment readiness and the significant role it plays in identifying whether an individual is likely to benefit from treatment (Casey et al., 2007; Day et al, 2009; Ward et al., 2004). Research indicates that a score below 72 on the CVTRQ indicates that an offender may not be ready for treatment (Casey et al., 2007). However, this study shows that when combined with moderate Antisocial Features, residents are less likely to have a successful outcome at Grendon if their score on the CVTRQ is below 78.5. Based on this study it is recommended that clinicians review the findings on the CVTRQ upon entry to Grendon and use it to help aid decision making and identify factors that may interfere with treatment. This study also provides some support for the MORM (Ward et al., 2004) as it indicates a relationship

between several internal factors such as self-esteem, relational style, and offence related guilt, and treatment outcome. Additional research could be conducted including both external factors, such as social milieu, as well as internal factors, to further investigate the applicability of the MORM within this environment.

It should also be noted that, although not included in the CRT due to missing data, the binary logistic regressions revealed a significant relationship between the presence of Histrionic Personality Disorder as identified by the PDQ, and outcome at HMP Grendon. The odds of a successful outcome at Grendon decreased by 72% given the presence of a histrionic personality disorder. To the author's knowledge, this is a novel finding in prison therapeutic communities. However, previous research has found a significant relationship between the presence of histrionic personality disorder and early attrition in a therapeutic community for the treatment of substance dependence (Samuel et al., 2011). As histrionic personality features are characterised by discomfort when the individual is not the centre of attention, excessive emotional reactions, and interpretation of relationships as more intimate than they are (American Psychiatric Association, 2022), this may cause significant difficulties within a therapeutic community. Personality disorders are not routinely screened for upon entry to Grendon, and do not form part of the eligibility criteria. However, as this study indicates that few people with histrionic personality features obtain a successful outcome at HMP Grendon, it may be helpful to explore the relevance of histrionic personality traits further. For example, to establish how and why individuals with such features have poorer outcomes in this provision and establish whether treatment could be adjusted accordingly to decrease the chances of a non- successful outcome.

Unfortunately, the model outlined in this analysis is not as accurate at predicting successful outcomes, as non-successful outcomes. The overall accuracy for successful outcomes is only 7.9% (Table 2.21). This may be due to the small sample size, as there are far fewer successful outcomes than non-successful outcomes and therefore the model has less power for predicting successful outcomes. It is also possible that successful outcomes are too heterogeneous, and so the model cannot predict many common profiles of success. However, the model can predict a successful outcome with 53.6% accuracy if an individual scores below 50.5 in Antisocial Features on the PAI. Having a low score in Antisocial Features was the only pathway for predicting a successful outcome at HMP Grendon. Previous research has found a positive relationship with antisocial personality factors and treatment non completion (Holdsworth et al., 2014; Pelissier et al., 2003), and our study supports this. As residents can be excluded from Grendon for displaying antisocial behaviour such as threats of violence, alcohol or drug use, it stands to reason that those lower in antisocial features may be more likely to withstand treatment and continue on to have a successful outcome at Grendon.

There is a lengthy history in the literature proposing a link between personality traits and offending behaviour (Eysenck, 1997; John et al., 1994; Van Dam et al., 2005). The basis of a therapeutic community is grounded in this theory. It is proposed that personality issues may have developed from difficult life experiences and can lead to criminal behaviour, but may be mitigated through intensive group-based therapy and a supportive social milieu (Newton et al., 1998). A therapeutic communities' goal is to promote long-term behavioural change by improving self-awareness, responsibility and interpersonal skills, and ultimately to reduce recidivism (Shuker & Sullivan, 2010). However, as our research has highlighted some personality traits are associated with unsuccessful outcomes at HMP Grendon, and

consequently we may infer that these traits could interfere with the residents being able to fully benefit from the treatment within a therapeutic community. As previously discussed, some of these personality traits such as low self-esteem and dominance may significantly impact the process within a therapeutic community. Our research has highlighted key areas for consideration when assessing suitability and treatment goals for an individual within a therapeutic community. It is suggested that the highlighted areas of personality associated with unsuccessful outcomes at HMP Grendon may also be appropriate targets for treatment within the community, if the individual is accepted for a place at Grendon. If treatment is adapted to work with these psychometric traits (e.g. Dominance, low Social Self-esteem etc.) then this may improve an individual's likelihood of remaining within the therapeutic community. This could consequently have a positive impact on the efficiency of the therapeutic community, and ability to reducing recidivism.

To ensure the validity of the CRT, future research should be conducted to cross-validate these decision rules on a dataset from which these decision rules were not derived. This model should therefore be applied to future residents who leave HMP Grendon to test the validity of the decision rules. This study has several clinical implications that should be considered. If cross-validated, this study could be used to adjust treatment to individual needs and improve the likelihood of a successful outcome at Grendon. At present, psychometric data is only reviewed when a resident has already been accepted into Grendon and a case formulation is conducted at approximately 6-9 months after entry. Based on our study, we would recommend that psychometric data is reviewed upon entry to Grendon to assess suitability to the therapeutic community and identify possible targets for treatment. Particular focus should be applied to profiles identified in Table 2.20 which may significantly interfere with

treatment. Further research should be conducted to assess the suitability and validity of necessary treatment adjustments.

To further research regarding successful outcomes at HMP Grendon, it may be worth considering if there are other predictors of success which have not been evaluated in this study. Previous research has highlighted that prison misconduct can significantly predict reoffending after release from prison (Cochran et al., 2014). Individuals with a record of violent misconduct were significantly more likely to reoffend (54%) when compared to those with non-violent misconduct (49%), and those who did not have a record of misconduct (42%). Additionally, research has shown that behavioural and educational prison programs which were associated with largest reductions in misconduct, were also associated with the largest reductions in reoffending after release from prison (French & Gendreau, 2006). Therefore, further research could focus on exploring if misconduct in prison is a more accurate predictor of success at Grendon than the psychometric variables used in this study.

Some limitations of this study need to be considered. Firstly, it should be noted that a very high bar for success was set for the purpose of this research. Success was defined as achieving all therapeutic goals and obtaining a progressive move on from Grendon. This allows us to look at the highest forms of successful outcome from HMP Grendon, and by including only those who have met all their therapeutic goals we can ensure that progressive moves can be well attributed to the therapeutic process. However, the limitation if this is that only 19% of residents achieve a successful outcome using this definition. The clinical regression model was not able to predict successful outcome with a high degree of accuracy using this definition. It is worth considering that this group may be heterogeneous in nature, and so the

model may not be able to predict common psychometric profiles of those who achieve successful outcomes. Our definition of success also excludes people who meet all therapeutic goals, but move from Grendon to a higher security prison due to their sentence type, or length of remaining sentence. It is possible that the model may be able to predict successful outcomes with higher levels of accuracy using an alternative definition of success. For example, it may be helpful to explore predictors of people who achieve a progressive move on from Grendon, regardless of whether they achieved all their therapeutic goals. If success was defined as gaining a progressive move on from Grendon, 29% of participants in the database would meet this criteria, which would substantially increase the sample size. It is suggested that predicting success is an important area for research which could allow us to better understand who is likely to benefit from attending Grendon. It is recommended that future research carefully considers alternative definitions of success to try and more accurately predict these outcomes.

A further limitation is that there are some difficulties with the data which may have inflated the number of people who appear to leave Grendon with a non-successful outcome. Our study includes some residents who were deemed unsuitable for Grendon during the assessment process, and were deselected from therapy at this early stage of the process. This subgroup is likely to be a substantially different population than those who were deemed suitable at assessment, but then went on to have a non-successful outcome. This therefore increases the heterogeneity of the non-successful population and may have inflated the number of non-successful outcomes. Grendon operates with an assessment unit to ensure that an individual's suitability for the therapeutic community is thoroughly assessed before they join one of the main wings. Although Grendon has always had some form of assessment unit, the assessment

process has been lengthened and enhanced since 2014. The assessment process is now 3-6 months, rather than the previous 2-3 months, and now closely replicates the main wings regime so that residents can adjust to the therapeutic community, and staff can accurately assess their suitability. Due to differences with data recording, the current study includes some residents who, prior to 2014, were deselected from therapy during the assessment process. For future research, it would be advisable to separate out data from individuals who were not deemed suitable to remain at Grendon after the assessment unit. This would provide more accurate statistics about non-successful outcomes and may allow for more accurate profile predictions. A CRT may then be able to predict profiles for who is likely to be deselected from therapy, who is deemed suitable for therapy, and whether they then go on to have a successful or non-successful outcome. Nevertheless, given our model's 97.6% accuracy in predicting non-successful outcomes, the inclusion of this subgroup in the present study could aid clinical decision-making during the assessment process. If clinicians interpreted psychometrics at this early stage of the process at Grendon, our model could then be used to aid decision-making as to someone's suitability for the therapeutic community, preparation needed before entering treatment, and identify targets for treatment.

What can other forensic therapeutic communities take away from this research?

This research provides initial evidence that individuals with some of the psychometric profiles identified in Table 2.20 may be significantly less likely to achieve a successful outcome within a UK prison therapeutic community. Such profiles include combinations of low Treatment Readiness as identified by the CVTRQ, Dominance, Antisocial Features, Suicidal

Ideation and Treatment Rejection as identified by the PAI, low Social Self-Esteem as identified by the CFSEI, and Confusion as identified by the PICT. There was some evidence that low Antisocial Features as identified by the PAI may be predictive of successful outcomes, however it should be noted that the model had limited accuracy for predicting successful outcomes. It is suggested that if this research is validated using another sample, then other therapeutic communities may benefit from administering and interpreting the CVTRQ, PAI, CFSEI and PICT within the assessment process. Due to the limited nature of this research and high consequence of exclusion, it is not suggested that the results of these psychometric evaluations should be used as exclusion criteria. However, they could be appropriately used to aid clinical decision-making about suitability for treatment, and highlight appropriate targets for therapy to improve chances of the resident obtaining a successful outcome after attending a DTC. To be most impactful, it is suggested that these psychometric evaluations are conducted and interpreted at the beginning of the assessment period, to allow clinicians to use them as part of their decision making about both suitability to attend a DTC, and inform treatment plans and targets if a resident is accepted into a DTC. There are of course limitations in applying this research to other DTCs which may significantly differ from HMP Grendon in terms of population, process and therapeutic regime. Appropriate caution should be taken when applying these results to other communities, and further research is required to further validate these findings.

Conclusion

In conclusion, the model outlined in this study provides us with a high level of confidence in predicting non-successful outcomes at Grendon. A variety of profiles were found to significantly increase chances of a non-successful outcome at Grendon, including combinations of Antisocial Features, Treatment Rejection, Confusion, Suicidal Ideation, Treatment Readiness, Dominance and Social Self-Esteem. The only pathway that was predictive of success was low levels of Antisocial Features. However, unfortunately the model was only able to predict successful outcomes with an overall accuracy of 7.9%. The current model may be useful for aiding clinical decision making about suitability to attend Grendon and identifying targets for treatment. Future research could use an adjusted definition of success, or explore alternative variables that may be predictive of successful outcome, such as prison conduct.

References

Alemohammad, M., Wood, J. L., Tapp, J., Moore, E., & Skelly, A. (2017). Support for the predictive validity of the multifactor offender readiness model (MORM): forensic patients' readiness and engagement with therapeutic groups. *Criminal Behaviour and Mental Health*, *27*(5), 421-442. https://doi.org/10.1002/cbm.2008

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV* (4th ed, text rev.). doi:10.1176/appi.books.9780890423349

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Arrindell, W. A., Hafkenscheid, A. J., & Emmelkamp, P. M. (1984). The Hostility and Direction of Hostility Questionnaire (HDHQ): A psychometric evaluation in psychiatric outpatients. *Personality and Individual Differences*, *5*(2), 221-231. https://doi.org/10.1016/0191-8869(84)90054-0

Battle, J. (1992). Culture-free self-esteem inventories: Examiner's manual (2nd ed.). PRO-ED.

Bender, D. S. (2005). The therapeutic alliance in the treatment of personality disorders. *Journal of Psychiatric Practice*, *11*(2), 73-87. https://doi.org/10 .1097/00131746-200503000- 00002

Bennett, P. (2007). Governing Grendon Prison's therapeutic communities: The big spin. In Parker, M. (Ed) *Dynamic Security: The Democratic Therapeutic Community in Prison* (pp. 203-212). Jessica Kingsley Publishers.

Bennett, J., & Shuker, R. (2017). The potential of prison-based democratic therapeutic communities. *International Journal of Prisoner Health*, *13*(1), 19-24. https://doi.org/10.1108/IJPH-08-2016-0036

Bennett, J., & Shuker, R. (2018). Hope, harmony and humanity: creating a positive social climate in a democratic therapeutic community prison and the implications for penal

practice. *Journal of Criminal Psychology*, *8*(1), 44-57. https://doi.org/10.1108/JCP-06-2017-0030

Birtchnell, J., Hammond, S., Horn, E., De Jong, C., & Kalaitzaki, A. (2013). A cross-national comparison of a shorter version of the person's relating to others questionnaire. *Clinical Psychology and Psychotherapy, 20*(1), 36–48. http://dx.doi.org/10.1002/cpp.789

Bridges, A. (2017). The contemporary pains of imprisonment: A study of HMP Grendon. *Howard League for Penal Reform*. MPhil thesis, University of Oxford. Accessed at: https://howardleague.org/wp-content/uploads/2019/01/The-contemporary-pains-ofimprisonment.pdf

Brookes, M. (2010). Putting principles into practice: the therapeutic community regime at HMP Grendon and its relationship with the 'good lives' model. In Shuker, R. & Sullivan, E (Eds.), *Grendon and the emergence of forensic therapeutic communities: Developments in research and practice*, (pp. 99-113). John Wiley & Sons. https://doi.org/10.1002/9780470661444.ch6

Busse, M., Whiteside, D., Waters, D., Hellings, J., & Ji, P. (2014). Exploring the reliability and component structure of the personality assessment inventory in a neuropsychological sample. *The Clinical Neuropsychologist*, *28*(2), 237-251. https://doi.org/10.1080/13854046.2013.876100

Caine, T. M., Foulds, G. A. & Hope, K. (1967) Manual of the Hostility–Direction of Hostility Questionnaire. London University Press.

Campbell, C., & Attwell, P. (2018). Intervening in secure settings. In Campbell, C. & Craissati, J (Eds.), *Managing Personality Disordered Offenders: A Pathways Approach* (115-144). Oxford University Press. https://doi.org/10.1093/med/9780198791874.003.0005

Casey S., Day A., Howells K., Ward T. (2007). Assessing suitability for offender rehabilitation: Development and validation of the Treatment Readiness Questionnaire.

Criminal Justice and Behaviour, 34, 1427-1440. https://doi.org/10.1177/0093854807305827

Cima, M., Merckelbach, H., Butt, C., Kremer, K., Knauer, E., & Schellbach-Matties, R. (2007). It was not me: Attribution of blame for criminal acts in psychiatric offenders. *Forensic Science International, 168,* 143–147. 10.1016/j.forsciint.2006.07.004

Cochran, J. C., Mears, D. P., Bales, W. D., & Stewart, E. A. (2014). Does inmate behaviour affect post-release offending? Investigating the misconduct-recidivism relationship among youth and adults. *Justice Quarterly*, *31*(6), 1044-1073. https://doi.org/10.1080/07418825.2012.736526

Cullen, E. (1994). Grendon: The therapeutic prison that works. *Therapeutic Communities. 15*(4), 301–311. https://psycnet.apa.org/record/1995-30842-001

Day, A., Howells, K., Casey, S., Ward, T., Chambers, J. C., & Birgden, A. (2009). Assessing treatment readiness in violent offenders. *Journal of Interpersonal Violence*, *24*(4), 618-635. https://doi.org/10.1177/0886260508317200

Dolan, R. (2017). HMP Grendon therapeutic community: The residents' perspective of the process of change. *Therapeutic Communities: The International Journal of Therapeutic Communities*, *38*(1), 23-31. https://doi.org/10.1108/TC-08-2016-0015

Daffern, M., Duggan, C., Huband, N., & Thomas, S. (2008). The impact of interpersonal style on aggression and treatment non-completion in patients with personality disorder admitted to a medium secure psychiatric unit. *Psychology, Crime & Law*, *14*(6), 481-492. https://doi.org/10.1080/10683160801948717

Eysenck, H. J. (1997). Personality and crime: Where do we stand? *Psychology, Crime & Law,* 2(3), 143–156. https://doi.org/10.1080/10683169608409773

Fox, S., De Koning, E., & Leicht, S. (2003). The relationship between attribution of blame for a violent act and EPQ-R sub-scales in male offenders with mental disorder. *Personality and Individual Differences, 34,* 1467–1475. https://doi.org/10.1016/S0191-8869(02)00128-9

French, S. A., & Gendreau, P. (2006). Reducing prison misconducts: What works!. *Criminal Justice and Behaviour*, *33*(2), 185-218. https://doi.org/10.1177/0093854805284406

Fuller, J., Tapp, J., & Draycott, S. (2019). Are guilt and shame in male forensic patients associated with treatment motivation and readiness? *Criminal Behaviour and Mental Health*, *29*(2), 111-121. https://doi.org/10.1002/cbm.2105

Gaab, S., Brazil, I. A., De Vries, M. G., & Bulten, B. H. (2020). The relationship between treatment alliance, social climate, and treatment readiness in long-term forensic psychiatric care: An explorative study. *International journal of offender therapy and comparative criminology*, *64*(9), 1013-1026. https://doi.org/10.1177/0306624X19899609

Genders, E., & Player, E. (1995). *Grendon: A study of a therapeutic prison.* Oxford University Press.

Gudjonsson, G. H., & Singh, K. K. (1989). The revised Gudjonsson blame attribution inventory. *Personality and Individual Differences, 10*(1), 67-70. https://doi.org/10.1016/0191-8869(89)90179-7

Holdsworth, E., Bowen, E., Brown, S., & Howat, D. (2014). Offender engagement in group programs and associations with offender characteristics and treatment factors: A review. *Aggression and Violent Behaviour*, *19*(2), 102–121. https://doi.org/10.1016/j.avb.2014.01.004

Howells, K., & Day, A. (2007). Readiness for treatment in high-risk offenders with personality disorders. *Psychology, Crime & Law*, *13*(1), 47-56. https://doi.org/10.1080/10683160600869767

Hyler, S. E. (1994). Personality Diagnostic Questionnaire–4 (PDQ–4). New York State Psychiatric Institute.

Jefferies, J. (2010). Psychodrama as part of core therapy at HMP Grendon. In Shuker, R. & Sullivan, E. (Eds.), *Grendon and the emergence of forensic therapeutic communities:*

Developments in research and practice, (pp. 137-151). John Wiley & Sons. https://onlinelibrary.wiley.com/doi/epdf/10.1002/9780470661444.ch8

John, O. P., Caspi, A., Robins, R. W., Moffitt, T. E., & Stouthamer-Loeber, M. (1994). The "little five": Exploring the nomological network of the five-factor model of personality in adolescent boys. *Child development,* 65(1), 160-178. https://doi.org/10.1111/j.1467-8624.1994.tb00742.x

Jones, L. (1997). Developing models for managing treatment integrity and efficacy in a prison-based TC: The Max Glatt Centre. In Cullen, E., Jones, L & Woodward, R (Eds.), *Therapeutic communities for offenders* (pp. 121-157). Wiley.

Marshall, P. (1997). A reconviction study of HMP Grendon therapeutic community. Home Office Research and Statistics Research Findings No. 53, Home Office, London. https://www.ojp.gov/ncjrs/virtual-library/abstracts/reconviction-study-hmp-grendon-therapeutic-community

McMurran, M., & Ward, T. (2010). Treatment readiness, treatment engagement and behaviour change. *Criminal Behaviour and Mental Health, 20*(2), 75-85. https://doi.org/10.1002/cbm.762

Miller, S., Brown, J., & Sees, C. (2004). A preliminary study identifying risk factors in drop-out from a prison therapeutic community. *Journal of Clinical Forensic Medicine, 11*(4), 189-197. https://doi.org/10.1016/j.jcfm.2003.12.003

Miller, S., Sees, C., and Brown, J. (2006). Key Aspects of Psychological Change in Residents of a Prison Therapeutic Community: A Focus Group Approach. *The Howard Journal, 45*(2), 116–128. https://doi.org/10.1111/j.1468-2311.2006.00409.x

Morey, L. C. (1991). Personality assessment inventory. Psychological Assessment Resources.

Morris, M. (2004). *Dangerous and Severe-Process, Programme and Person: Grendon's Work.* Jessica Kingsley Publishers.

Newton, M. (1998). Changes in measures of personality, hostility, and locus of control during residence in a prison therapeutic community. *Legal and Criminological Psychology, 3*(2), 209-223. https://doi.org/10.1111/j.2044-8333.1998.tb00362.x

Pelissier, B., Camp, S. D. & Motivans, M. (2003). Staying in Treatment. *Psychology of Addictive Behaviours, 17*(2), 134-141. https://doi.org/10.1037/0893-164X.17.2.134

Pointon, V. J., & Roberts, R. (2023). Psychometric attributes associated with attrition within a prison-based democratic therapeutic community. *Therapeutic Communities: The International Journal of Therapeutic Communities,* 44(1), 6-16.

Rawlings, B., & Haigh, R. (2017). Therapeutic communities and planned environments for serious offenders in English prisons. *The British Journal of Psychiatry Advances, 23*(5), 338-346. doi:10.1192/apt.bp.115.015636

Samstag, L.W., Batchelder, S., Muran, J.C., et al. (1998). Predicting treatment failure from in-session interpersonal variables. *The Journal of Psychotherapy Practice and Research, (5)*126–43

Samuel, D. B., LaPaglia, D. M., Maccarelli, L. M., Moore, B. A., & Ball, S. A. (2011). Personality disorders and retention in a therapeutic community for substance dependence. *The American Journal on Addictions*, *20*(6), 555-562. https://doi.org/10.1111/j.1521-0391.2011.00174.x

Schanz, Y. Y. (2017). The relationship between aggression and self-esteem among prison inmates. *Journal of Aggression, Maltreatment & Trauma, 26*(9), 989-1005. https://doi.org/10.1080/10926771.2017.1331940

Shine, J. (2001). Characteristics of inmates admitted to Grendon therapeutic prison and their relationships to length of stay. *International Journal of Offender Therapy and Comparative Criminology, 45*(2), 252-265. https://doi.org/10.1177/0306624X01452010

Shuker, R., & Newton, M. (2008). Treatment outcome following intervention in a prison-based therapeutic community: a study of the relationship between reduction in

criminogenic risk and improved psychological well-being. *The British Journal of Forensic Practice, 10*(3), 33-44. https://doi.org/10.1108/14636646200800018

Shuker, R., & Sullivan, E. (Eds.). (2010). Grendon and the emergence of forensic therapeutic communities: Developments in research and practice. John Wiley & Sons.

Skett, S. & Lewis, C. (2019), Development of the offender personality disorder pathway: a summary of the underpinning evidence, *Probation Journal*, 66*(2),*167-180. https://doi.org/10.1177/0264550519832370

Taylor, R. (2000). A seven-year reconviction study of HMP Grendon therapeutic community. Home Office Research, Development and Statistics Directorate. https://www.ojp.gov/ncjrs/virtual-library/abstracts/seven-year-reconviction-study-hmp-grendon-therapeutic-community

Tollinton, H. P. (1966). Grendon Prison. *British Journal of Criminology*, *6*(39).

Valentiner, D. P., Skowronski, J. J., McGrath, P. B., Smith, S. A., & Renner, K. A. (2011). Self-verification and social anxiety: preference for negative social feedback and low social self-esteem. *Behavioural and Cognitive Psychotherapy*, *39*(5), 601-617. doi:10.1017/S1352465811000300

Van Dam, C., Janssens, J. M., & De Bruyn, E. E. (2005). PEN, Big Five, juvenile delinquency and criminal recidivism. *Personality and individual differences*, 39(1), 7-19. https://doi.org/10.1016/j.paid.2004.06.016

Walters, G. D. (1995). The psychological inventory of criminal thinking styles: Part I: Reliability and preliminary validity. *Criminal Justice and Behaviour*, *22*(3), 307-325. https://doi.org/10.1177/0093854895022003008

Ward, T., Day, A., Howells, K., & Birgden, A. (2004). The multifactor offender readiness model. *Aggression and Violent Behaviour,* 9, 645-673. https://doi.org/10.1016/j.avb.2003.08.001

Watson, R., Thomas, S., & Daffern, M. (2017). The impact of interpersonal style on ruptures and repairs in the therapeutic alliance between offenders and therapists in sex offender treatment. *Sexual Abuse*, *29*(7), 709-728. https://doi.org/10.1177/1079063215617514

Weizmann-Henelius, G., Sailas, E., Viemerö, V., & Eronen, M. (2002). Violent women, blame attribution, crime, and personality. *Psychopathology,* 35, 355–361. https://doi.org/10.1159/000068590

Wylie, B. (2010). Self and social function: Art therapy and readiness for treatment in a therapeutic community prison. In Shuker, R. & Sullivan, E. (Eds.), *Grendon and the Emergence of Forensic Therapeutic Communities: Developments in Research and Practice*, (153-169). John Wiley & Sons. https://doi.org/10.1002/9780470661444.ch9

Zeigler-Hill, V. (2013). *Self-esteem*. Psychology Press.

Chapter 3: Press release for the meta-analysis

University of Birmingham News Release

Embargo: TBC

Women in the prison services across the world experience disproportionate rates of Post Traumatic Stress Disorder, compared to men

Nearly one quarter of women in the prison systems across the world have post traumatic stress disorder (PTSD), compared with only 6% of men in prison.

Psychologists at The University of Birmingham have conducted a large international review of existing research into PTSD in female prison populations. The review included 7266 participants, across 11 countries worldwide. The research concludes that 24% of women in prison suffer from PTSD, which demonstrates the shocking level of poor mental health in prison systems across the globe. This can be compared to previous research that found 6% of men in prison experience PTSD (Baranyi et al., 2018).

PTSD can develop after exposure to traumatic events, and is characterised by symptoms such as flashbacks, nightmares, avoidance, excessive attention to potential threat, physical symptoms such as feeling sick or sweating, and sleep problems. Women in the prison service are more likely to have experienced traumatic events in their childhood, such as sexual abuse

and neglect (Byrne & Howells, 2000). Such exposure to traumatic events can often lead to the development of PTSD.

Previous research shows that PTSD can prevent people from benefitting from rehabilitative programmes in prison, and is associated with higher rates of reoffending (Allely & Allely, 2020; Karatizias et al., 2018; Kubiak, 2004). If left untreated, women who are experiencing PTSD are more likely to reoffend, and subsequently further harm can be done to society and potential victims. However, PTSD is a treatable condition, and therapy such as Eye Movement Desensitisation and Reprocessing and Trauma Focused- Cognitive Behavioural Therapy are widely used and effective treatments (Seidler et al., 2006). Research into PTSD treatments in prison is limited, however initial studies are showing encouraging results (Malik et al., 2023). Research indicates that increasing rates of diagnosis and treatment of PTSD in prisons could reduce offending rates (Sadeh & McNiel, 2015) and so it is important that we understand the scale of the problem. If prisons offered more treatment for conditions such as PTSD, we may substantially reduce reoffending rates, and consequently this would have a beneficial effect for wider society.

“These staggering statistics reveal the high level of women experiencing PTSD in prison services across the world. More needs to be done to identify and treat this condition effectively so women can benefit from rehabilitative programmes in prison, and reduce the risk of reoffending again in future” says Annie Sillence, Trainee Clinical Psychologist at the University of Birmingham.

This important research highlights the need for criminal justice systems across the world to think about the disproportionate rates of women with PTSD who are being incarcerated. Here in the UK, the government's Female Offender Strategy has highlighted its ambition for the widespread implementation of a trauma informed approach in all women's prisons (Ministry of Justice, 2018). This includes increased training for staff, and prisoner-led trauma support groups. These adaptations for women experiencing PTSD could be vital to ensure they are able to better able to benefit from rehabilitation in prison, and reduce the risk of future offending. Further research is required to investigate to what extent these changes are being implemented, and how effective they are.

This research was funded by Health Education England's Clinical Psychology Doctoral Programme.

ENDS

For media enquiries please contact Beck Lockwood, Press Office, University of Birmingham, tel: email:

Notes to editor:

The University of Birmingham is ranked amongst the world's top 100 institutions. Its work brings people from across the world to Birmingham, including researchers, teachers and more than 8,000 international students from over 150 countries.

References

Allely, C. S., & Allely, B. (2020). Post traumatic stress disorder in incarcerated populations: Current clinical considerations and recommendations. *Journal of Criminal Psychology*, *10*(1), 30–42. https://doi.org/10.1108/JCP-09-2019-0044

Baranyi, G., Cassidy, M., Fazel, S., Priebe, S., & Mundt, A. P. (2018). Prevalence of posttraumatic stress disorder in prisoners. *Epidemiologic reviews*, *40*(1), 134-145. https://doi.org/10.1093/epirev/mxx015

Karatzias, T., Power, K., Woolston, C., Apurva, P., Begley, A., Mirza, K., Conway, L., Quinn, C., Jowett, S., Howard, R., & Purdie, A. (2018). Multiple traumatic experiences, post-traumatic stress disorder and offending behaviour in female prisoners. *Criminal Behaviour and Mental Health*, *28*(1), 72–84. https://doi.org/10.1002/cbm.2043

Kubiak, S. P. (2004). The Effects of PTSD on Treatment Adherence, Drug Relapse, and Criminal Recidivism in a Sample of Incarcerated Men and Women. *Research on Social Work Practice*, *14*(6), 424–433. https://doi.org/10.1177/1049731504265837

Malik, N., Facer-Irwin, E., Dickson, H., Bird, A., & MacManus, D. (2023). The Effectiveness of Trauma-Focused Interventions in Prison Settings: A Systematic Review and Meta-Analysis. *Trauma, Violence, & Abuse*, *24*(2), 844–857. https://doi.org/10.1177/15248380211043890

Ministry of Justice. (2018). Female offender strategy. London: HMSO

Sadeh, N., & McNiel, D. E. (2015). Posttraumatic Stress Disorder Increases Risk of Criminal Recidivism Among Justice-Involved Persons With Mental Disorders. *Criminal Justice and Behaviour, 42*(6), 573–586. https://doi.org/10.1177/0093854814556880

Seidler, G. H., & Wagner, F. E. (2006). Comparing the efficacy of EMDR and trauma-focused cognitive-behavioural therapy in the treatment of PTSD: a meta-analytic study. *Psychological medicine*, *36*(11), 1515-1522. doi:10.1017/S0033291706007963

Chapter 4: Press release for the empirical research paper

University of Birmingham News Release

Embargo: TBC

Factors predicting non-successful outcomes at a therapeutic prison discovered

University of Birmingham researchers have discovered personality factors that predict non-successful outcomes at HMP Grendon, the UK's leading therapeutic prison.

HMP Grendon is a therapeutic prison in Buckinghamshire for men who have committed serious and violent offences. It is a revolutionary prison and aims to reduce reoffending by treating men with group therapy. It works with some of the most dangerous offenders in the prison system, and previous research has shown that it significantly reduces reoffending, when compared with men who attend mainstream prisons (Taylor, 2000). It plays an important role in our prison service and helps keep our society safer. However, HMP Grendon does not work well for everyone, and little is known about who is likely to have a successful outcome. Some men are deemed unsuitable for the treatment and are transferred to other prisons. Valuable resources may be wasted by treating men ineffectively, and so there is huge cost and societal benefit to further increasing the efficiency of this vital facility.

“This research is really exciting, if we can understand who doesn’t get a successful outcome at HMP Grendon, we may be able to adapt the treatment to better suit the individual’s needs, and reduce reoffending rates” says Annie Sillence, a Trainee Clinical Psychologist at the University of Birmingham.

Researchers at the University of Birmingham conducted analysis with men who have attended HMP Grendon over the last 30 years. They investigated the relationship between certain personality factors, and successful and non-successful outcomes at HMP Grendon. Successful outcomes included were meeting goals set during therapy, and moving on to lower security prisons. The researchers discovered certain personality factors, such as antisocial features and dominance, could predict non-successful outcomes at HMP Grendon with a 98.1% accuracy. They also identified some factors such as social self-esteem which can improve the chances of a successful outcome at HMP Grendon.

“Men at HMP Grendon are some of the most traumatised men in the prison service. As well as being the perpetrators of crime, they have experienced the most unimaginable suffering in their childhoods” reports Annie Sillence.

If we can better understand factors that will help improve chances of engaging with treatment at HMP Grendon successfully, we can reduce the risk of these offenders committing more crime again in future. By improving social self-esteem, research indicates that we can increase the chance of men gaining a successful outcome at HMP Grendon, and they are therefore significantly less likely to offend again in future.

A place at HMP Grendon has been demonstrated to be a cost-effective way of rehabilitating prisoners (Albertson, 2013). This is due to its reductions in reoffending and diversion from other higher cost facilities such as high security prisons and mental health hospitals. Therefore, research such as this that may help increase the efficiency of the prison would also reduce the cost to the taxpayer, as well as reducing the amount of future victims of crime.

As well as HMP Grendon, there are other therapeutic communities across the country that work within prisons, and aim to reduce reoffending with group therapy. This research could be of vital importance to not only HMP Grendon, but also to other therapeutic communities. If assessed upon their arrival to a therapeutic community, we could ensure that these personality factors that interfere with treatment are identified, and adjust their therapy accordingly. Researchers hope that this could reduce dropout rates at therapeutic communities, and increase the number of people who benefit from the service.

Further research has been recommended to better understand who succeeds at HMP Grendon, as well as identifying those who do not currently succeed. This may help the prison further improve the prisons efficiency, and bring benefit to wider society by rehabilitating offenders.

This research was funded by Health Education England's Clinical Psychology Doctoral Programme.

ENDS

For media enquiries please contact Beck Lockwood, Press Office, University of Birmingham, tel: +44 email:

Notes to editor:

The University of Birmingham is ranked amongst the world's top 100 institutions. Its work brings people from across the world to Birmingham, including researchers, teachers and more than 8,000 international students from over 150 countries.

References

Albertson, K. (2013), “HMP Grendon: a first approximation cost/benefit analysis”, unpublished report, Manchester Metropolitan University, Manchester.

Taylor, R. (2000). A seven-year reconviction study of HMP Grendon therapeutic community. London: Home Office Research, Development and Statistics Directorate. https://www.ojp.gov/ncjrs/virtual-library/abstracts/seven-year-reconviction-study-hmp-grendon-therapeutic-community

Appendices

Appendix 1: Ethical approval from University of Birmingham

UNIVERSITY OF BIRMINGHAM

Dear Christopher Jones,

RE: Exploratory analysis of predictors of treatment failure at HMP Grendon

Reference: Ethical Review Form

Thank you for your application for ethical review for the above project, which has now been reviewed by the Science, Technology, Engineering and Mathematics Committee. On behalf of the Committee, I am pleased to confirm ethics approval for your project, subject to your adherence to the following conditions:

- Please ensure all relevant external approvals are in place at HMP Grendon prior to commencing the project.

For clarification, as long as the conditions above are met and the details of the proposed work do not change, your project has ethics approval and no further action is necessary.

Any adverse events occurring during the study should be promptly brought to the Committee's attention by the Principal Investigator and may necessitate further ethical review.

Please ensure that the relevant requirements within the University's Code of Practice for Research and the information and guidance provided on the University's ethics webpages (available at https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Links-and-Resources.aspx) are adhered to.

Please be aware that whilst Health and Safety (H&S) issues may be considered during the ethical review process, you are still required to follow the University's guidance on H&S and to ensure that H&S risk assessments have been carried out as appropriate. For further information about this, please contact your School H&S representative or the University's H&S Unit at healthandsafety@contacts.bham.ac.uk.

Kind regards,

The Co-Chairs of the Science, Technology, Engineering and Mathematics Committee

E-mail: ethics-queries@contacts.bham.ac.uk

Appendix 2: Ethical approval from University of Birmingham after amendment

UNIVERSITY OF BIRMINGHAM

Dear Mr Chris Jone, Annie Sillence

RE: Exploratory analysis of predictors of treatment failure at HMP Grendon

Application for Ethical Amendment: ERN_0498-Sep2023

Thank you for your application for amendment to the above project, which was reviewed by the Science, Technology, Engineering and Mathematics committee.

On behalf of the Committee, I confirm that this amendment has full ethical approval.

Any adverse events occurring during the study should be promptly brought to the Committee's attention by the Principal Investigator and may necessitate further ethical review.

Please ensure that the relevant requirements within the University's Code of Practice for Research and the information and guidance provided on the University's ethics webpages (available at https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Links-and-Resources.aspx) are adhered to.

Please be aware that whilst Health and Safety (H&S) issues may be considered during the ethical review process, you are still required to follow the University's guidance on H&S and to ensure that H&S risk assessments have been carried out as appropriate. For further information about this, please contact your School H&S representative or the University's H&S Unit at healthandsafety@contacts.bham.ac.uk.

Kind regards,

The Co-Chairs of the Science, Technology, Engineering and Mathematics Committee

E-mail: ethics-queries@contacts.bham.ac.uk

Appendix 3: Measures

Measures used in the empirical research are included in this appendix only if they are freely available. Most measures are not included due to copyright laws.

Gudjonsson Blame Attribution Inventory

GBAI-R

Name:

Date:

Nature of the Offence:

Below are a number of statements related to the crime(s) you committed. Please read each item carefully and decide whether the statement is TRUE or FALSE as it applies to you personally. If the statement is true as applies to you then circle True; and if it is false as applied to you then circle False.

1.	I feel very ashamed of the crime(s) I committed.	True	False
2.	I am entirely to blame for my crime(s).	True	False
3.	I did not deserve to get caught for the crime(s) I committed.	True	False
4.	I am constantly troubled by my conscience for the crime(s) I committed.	True	False
5.	I will never forgive myself for the crime(s) I committed.	True	False
6.	I feel no remorse or guilt for the crime(s) I committed.	True	False
7.	I am responsible for my criminal act(s).	True	False
8.	It is definitely not in my nature to commit crimes.	True	False
9.	I should not blame myself for the crime(s) I committed.	True	False
10.	At the time of the crime(s) I was fully aware of what I was doing.	True	False
11.	I would not have committed the crime(s) I did if I had not lost control of myself.	True	False
12.	I should not blame other people for my crime(s).	True	False
13.	The crime(s) I committed was very much out of character.	True	False
14.	I hate myself for the crime(s) I committed.	True	False
15.	Society is to blame for the crime(s) I committed.	True	False
16.	I should not be punished for what I did.	True	False
17.	I was feeling no different to usual at the time of the crime(s).	True	False

18.	In my case the victim(s) was largely to blame for my crime(s).	True	False
19.	I would not have committed any crime(s) if I had not been seriously provoked by the victim(s)/society	True	False
20.	What I did was beyond my control.	True	False
21.	I deserved to be caught for what I did.	True	False
22.	I would have been better off if I had not been caught.	True	False
23.	I constantly have the urge to punish myself for the crime(s) I committed.	True	False
24.	I fear that people will never accept me because of the crime(s) I committed.	True	False
25.	I was very depressed when I committed the crime(s).	True	False
26.	I was in no way provoked into committing a crime.	True	False
27.	I have no need to feel ashamed of what I did.	True	False
28.	I feel annoyed that I was caught.	True	False
29.	I must have been crazy to commit the crime(s) I did.	True	False
30.	There is no such thing as an innocent victim in my case.	True	False
31.	Other people are to blame for my crime(s).	True	False
32.	I could have avoided getting into trouble.	True	False
33.	I had very good reasons for committing the crime(s) I did.	True	False
34.	I should not punish myself for what I did.	True	False
35.	I deserve to be severely punished for the crime(s) I committed.	True	False
36.	I would certainly not have committed the crime(s) I did if I had been mentally well.	True	False
37.	I have no serious regrets about what I did.	True	False
38.	I was under a great deal of stress/pressure when I committed the crime(s).	True	False
39.	I would very much like to make amends for what I did.	True	False
40.	I sometimes have nightmares about the crime(s) I committed.	True	False
41.	I was in full control of my actions.	True	False
42.	I have no excuse for the crime(s) I committed.	True	False

Scoring:

M	E	G

Appendix 4: Consent statement signed by all residents at HMP Grendon

"As part of your decision to engage in treatment here at HMP Grendon, you are volunteering to engage fully in the therapeutic regime of the prison in order to address your areas of risk and need.

In order to help with treatment planning and assess your progress in treatment, you will be asked to provide information about yourself and to complete a number of psychological interviews, and questionnaires. These are used to help with identifying areas that you will need to work on. This information will be shared with other members of staff and maybe used to inform decisions about my progress through sentence and release.

In order to improve the regime and the treatment provided, information from assessments will also be collated and used in confidence by the research and development unit or, approved external researchers. I acknowledge that if information from assessments is used for research I will not be identified"

Appendix 5: External approval letter from HMP Grendon

Letter of Confirmation

To Whom It May Concern

I hereby confirm that Mrs Annie Sillence's research, titled "*Exploratory analysis of predictors of treatment failure at HMP Grendon*" has been approved and will be supported by HMP Grendon.

Yours sincerely,

XXXXXX

Head of Clinical Services

HMP Grendon

Appendix 6: University of Birmingham, Centre for Applied Psychology - Meta-Analysis Strategy

Handling of data that violates analysis assumptions

Transformation of effects for calculations and back transformation for presentation

The event rates and relative risk estimates in primary study were log transformed prior to numerical synthesis however, unless otherwise indicated, the values presented in tables and figures have been back-transformed to their original format for clarity of presentation.

Missing data and zero frequency data

Event rates with a zero count can cause numerical problems in the synthesis of relative risk and event rates. Zero counts typically occur in small studies in which the sample size does not afford accurate estimation of the true event rate (i.e., were the zero-event rate reflects the lack of opportunity to observe an event rather than a true absence of events). Studies with a sample size less than XX participants were excluded from this review as unable to estimate the true event rate. If a study had a sample size greater than forty but an event rates equal to zero then a small constant was added (i.e., 0.5) was added to the zero-event rate to avoid division by zero errors.

Normalisation and variance stabilisation

The DerSimonian and Laird method (DerSimonian & Laird, 1986) is the simplest and most commonly used method for calculating the between studies variation (tau) for fitting the random effects model. However, the DerSimonian and Laird method assumes that the random effect is normally distributed in the population and therefore the effects sizes reported in the primary studies should also approximate a normal distribution. The log transformation of the event rates and relative risk estimates serve to normalise the distribution of effects and stabilise the variance of the estimates prior to synthesis using the DerSimonian and Laird method.

Please note that the

- log transformation is used for event rates and relative risk (or risk difference) effects,
- Fisher's Z transformation for correlations,
- log transformation is used for phi coefficients,
- log transformation is used mean difference that show markedly non-normal distributions (although it is better to use the Restricted maximum-likelihood estimator (ReML) to estimate tau from the raw difference scores).
- Cohen's *d* has been shown to systematically overestimate the absolute value of the SMD in small samples (Borenstein, 2009). This bias can be removed by transforming Cohen's *d* into an unbiased

estimate known as Hedge's *g* (Hedges, 1981) for the calculations and back transforming into Cohen's *d* for interpretation and reporting in tables and figures.

For meta-analysis of alpha coefficient then the transformation reported by Bonett (2002, 2010) is most commonly used to correct for issues relating to normalisation and variance stabilisation.

If you have selected other than the DerSimonian and Laird method for calculating the random effects model then you should provide an explanation for your choice (in the section on the Omnibus test) and discuss if transformation is required.

The omnibus test

The omnibus test can be calculated using either the fixed effects or the random effects models. Under the fixed-effect model we assume that the true effect size for all studies is identical, and the only reason the effect size varies between studies is sampling error (error in estimating the effect size). Therefore, when assigning weights to the different studies we can largely ignore the information in the smaller studies since we have better information about the same effect size in the larger studies. It makes sense to use the fixed-effect model if two conditions are met. First, we believe that all the studies included in the analysis are functionally identical (i.e., all studies have a uniformly excellent methodology). Secondly, our goal is to compute the common effect size for the identified population, and not to generalize to other populations. In point of fact, effects in psychological studies are likely to vary as a result of a number of uncontrolled factors (e.g., the distribution methodological weakness across studies, uncontrolled moderators, natural variation in the effect that is being measured).

In contrast, under the random-effects model the goal is not to estimate one true effect but to estimate the mean of a distribution of possible effects (which may show true variation due to the idiosyncratic characteristics of the individual or the unique circumstances of the intervention or exposure). Since each study provides information about a different effect size, we want to be sure that all these effect sizes are represented in the summary estimate. This means that we cannot discount a small study by giving it a very small weight (the way we would in a fixed-effect analysis). The estimate provided by that study may be imprecise, but it is information about an effect that no other study has estimated. By the same logic we cannot give too much weight to a very large study (the way we might in a fixed-effect analysis). Our goal is to estimate the mean effect in a range of studies, and we do not want that overall estimate to be overly influenced by any one of them. When the researcher has gathered data from studies that had been undertaken by researchers operating independently (and will therefore show different methodological strengths and weaknesses), it would be unlikely that all the studies are functionally equivalent. Typically, the participants and/or interventions in these studies would have differed in ways that would have impacted on the results, and therefore we should not assume a common effect size. Therefore, in in the case of the current review the random-effects model is more easily justified than the fixed-effect model.

The DerSimonian and Laird method is the simplest and most commonly used method for calculating the between studies variation (tau) for fitting the random effects model.

If you have selected other than the DerSimonian and Laird method for calculating the random effects model then you should provide an explanation for your choice. The list below may be of assistance.

1. 'DL' is the and it is the default for effects that are considered to be normally distributed in the population.
2. Sidik-Jonkman estimator ('SJ") is known to perform better than the DerSimonian-Laird estimate when trials of similar size are combined.
3. 'ReML' is the Restricted maximum-likelihood estimator and 'ML' is the Maximum-likelihood estimator. These estimators are more robust than the DL estimator to non-normal distributions of effects. Usually, the Restricted Maximum-likelihood estimator should be used in preference to the Maximum-likelihood estimator. The Maximum-likelihood estimator is provided in case you wish to compare with the results of a previous meta-analysis.
4. Hunter-Schmidt estimator ('HS') tends to over-estimate the amount of variance due to sampling error and is included only for the purpose of comparison with a previous meta-analysis.
5. Hedges estimator ("HE") is used to estimate heterogeneity in Cochrane meta-analyses and is included for comparison to such reviews.
6. Empirical Bayes estimator ("EB") can be used to identify whether new trial evidence changes the conclusion of a previous review. BEWARE: To implement this procedure, you will need the help of your research tutor.

Handling problematic variance

Defining problematic variance

An effect is considered heterogeneous if it presents with variation from the meta-analysis synthesis that cannot be attributed to true variation in the distribution population effect. Heterogeneity can result from methodological variation in the studies, measurement error or uncontrolled individual difference factors within the body of literature. Higgins I^2 is a commonly used measure of heterogeneity, with greater values of I^2 indicating variation in effect that cannot be attributed to true variation in the distribution of effect in the population. As there is considerable variation in methodologies of the primary studies that was used to calculate the meta-analytic synthesis, problematic heterogeneity was defined as a Higgins I^2 value greater than 75%.

Estimation of unexplained variance due to methodological factors and uncontrolled covariates

If problematic heterogeneity is observed then a leave-one-out analysis will be conducted to identify primary studies that exert a disproportionately influential effect on the meta-analytic synthesis. Any such study will be reviewed with regard to the possibility of exclusion due to risk of bias.

In addition, subgroup analyses and meta regression will be used to attempt to identify the source or sources of problematic heterogeneity and the attenuated estimate of the synthesis will be reported.

The quality effects model

In the random effects model the precision of an effect is usually estimated as a function of the sample size from which the effect is derived. The quality effects model (Doi & Thalib, 2008) extends the random effects model by explicitly including rating of methodological quality in addition to the size of the sample in the estimation of precision. In this review the quality effects model was calculated using the total score from the risk of bias ratings reported in section XX. The quality effects model can be interpreted as the meta-analytic synthesise that would have been obtained had all of the studies been of the same methodological quality as the best study in the review. Accordingly, the quality effects model provides a measure of attrition attributable to methodological variation.

Identifying Influential Studies

To examine whether any particular study or studies are exerting a disproportionately high influence on the overall meta-analytic effect, a “one left out” procedure was conducted. This procedure identifies individual studies with a disproportionate influence on the quantitative synthesis, by observing the impact of removing each study in turn. If omitting a study results in an effect that lies outside of the 95% CI for the complete meta-analysis then that study is deemed to have a disproportionate influence and is remove from the omnibus test.

Identifying Publication Bias and Small Study Effects

For outcomes with a sufficient number of primary studies, publication bias and small study effects will be identified through visual and statistical inspection of the funnel plot. A funnel plot is a scatterplot of the effects from against a measure of study precision. It is used primarily as a visual aid for detecting systematic heterogeneity.

In the absence of publication bias, it is assumed that studies with high precision will be plotted near the average (i.e., the meta-analytic synthesis), and studies with low precision will be spread evenly on both sides of the average, creating a roughly funnel-shaped distribution where the distance from the average is inversely proportionate to the precision of the study. A symmetric inverted funnel shape arises from a 'well-behaved' data set, in which publication bias is unlikely whereas deviation from this shape can indicate publication bias especially if there is an absence of studies in the region associated with small samples sizes and non-significant effects.

If publication bias is identified, then a trim and fill procedure (Duval & Tweedle, 2000a; Duval & Tweedle, 2000b) will be undertaken. The trim and fill procedure builds on the assumption that publication bias would lead to an asymmetrical funnel plot. Trim and fill procedure uses an iterative algorythm to remove the most extreme small studies from the side of the funnel plot associated with positive effects, re-computing the effect size at each iteration until the funnel plot is symmetric about the (corrected) effect size. In theory, this will yield an unbiased estimate of the effect size. While this trimming yields the adjusted effect size, it also reduces the variance of the effects, yielding a too narrow confidence interval. Therefore, the algorithm then adds the original studies back into the analysis, and imputes a mirror image for each on the side of the funnel plot associated with negative effects.

In addition, the fail-safe N will also be calculated (Rosenthal, 1979). The fail-safe N is an estimation of the number of missing studies that would need to be retrieved for the effect to be no longer significant. If this number is large (relative to the number of primary studies in the meta-analysis) then the omnibus test can be considered robust to the effects of publication bias.

Planned Contrasts

Where specific a priori hypothesis made been posited, then sub-group analysis will be conducted for categorical moderators and meta-regression will be calculated for continuous moderators.

Analysis of Sub-groups

Where categorical moderators are considered then summary effects and associated heterogeneity measures will be calculated for each of the sub-groups. The significance of the difference between the sub-groups will be evaluated by comparison of their 95% confidence intervals.

Potential moderators of the effect will be explored using a series of subgroup analyses. The significance of sub-group differences will be evaluated using the Q statistic, which may be view as an extension of analysis of variance. The Q statistic is calculated by summing the within-studies variation (the weighted sum of squares of all of the studies within a subgroup about the mean of the subgroup) across all subgroups and the subtracting this from the total variance (i.e., the weighted sum of squares between all of the studies and the overall grand mean). The resulting Q statistic therefore represents the ., the weighted sum of squares attributable to between studies variation and conforms to a chi-squared distribution(Borenstein, 2009). A 95% confidence interval for each subgroup will be used to determine the significance of the pairwise differences between the sub-groups.

Meta-Regression

Meta-regression is similar to simple regression, in that the effects of the primary studies are predicted according to the values of one or more explanatory variables. However,

larger studies have more influence on the relationship than smaller studies, since studies are weighted by the precision of their respective effect estimate. The explanatory variables are typically characteristics of studies or participants that might influence the size of intervention effect.