Prolonging Release in High-Risk, Complex, Violent, Male Offenders

By Darius Cavina

Thesis Submitted to the College of Life and Environmental Sciences, School of Psychology at the University of Birmingham for the Continued Professional Development route to the Doctorate in Forensic Psychology Practice (ForenPsyD)

NOVEMBER 2023

Abstract

The treatment and management of high-risk, violent, complex men released from custody to the community has gained much attention in recent years. This group are disproportionately affected by traits of personality disorder, making the processes of risk management and treatment both complex and slow. This thesis provides a detailed exploration of who services should target for intensive community interventions, what types of intervention appear to work with this complex group, and how best to measure the efficacy of such interventions. A general introduction to this area of research is presented, followed by a systematic review of the literature on how this group have typically been managed in the community. The findings are mixed, with some studies showing no effect of community interventions using recidivism outcome measures, but several high-quality studies finding strong intervention effects when contact begins in custody and continues into the community, thus representing a transitional service. Original research is then presented examining the efficacy of existing Intensive Intervention and Risk Management Services (IIRMS) and the Intensive Response Intensive Service (IRiS) project. Both cohorts are compared with large control groups matched using key variables. Whilst findings did not establish an overall intervention effect on survival time, for all participants, higher risk and personality complexity scores were predictive of shorter survival times in the community. The problems of defining and measuring 'success in the community' are discussed and a detailed examination of the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) is presented. Discussions about the strengths and weaknesses of using recidivism measures, and the need to incorporate more nuanced measures of progress, including self-report well-being measures, are presented. Finally, a general discussion brings together the findings from this thesis, making recommendations for practice, so that those most in need are offered high-quality services that can be effectively measured.

Acknowledgments

I would like to thank my supervisor, Dr Caroline Oliver, for her guidance and support throughout this process. Not only has she provided invaluable feedback and sage advice, but she has also been a critical friend when needed and kept me motivated through difficult months when data access was challenging.

I am also grateful to the group of professionals who explored numerous different avenues before succeeding in making high-quality, large-scale data available for my research. They helped to mitigate the difficulties of accessing data in a global pandemic. Thank you to Kristy Amour for accessing and tidying up this data, to Kimberlee Thersby for providing the OPDP screening data, to Dr Carine Lewis for supporting my research from the start and to Aisling O'Meara for answering all my questions in double quick time, despite completing her own Doctorate at the same time.

Substantial thanks go to the five Intensive Intervention and Risk Management Services that agreed to release their data for my research. I am also grateful to the IRiS team in Bristol who were extremely supportive and committed to the inclusion of their data in my research.

Analysis of the data was complex and only possible with the support of a group of professionals who not only understood the statistics involved but also the nature of the data being used. I am extremely grateful to Aidan Mews for his time, knowledge and advice regarding PSM in particular. I am also grateful to Dr Chris Jones for his generous offer of time and extremely helpful feedback.

Without the support and flexibility of my colleagues and managers within the Wales Offender Personality Disorder Pathway team, this thesis would never have been finished. Not only did they support me, but they encouraged me and constantly reminded me of the importance of this work. Thank you very much.

Finally, and most importantly, I owe a debt of gratitude to my family, who have toiled as much as I have. You have allowed me to explore an area of research about which I am passionate. I am grateful beyond words – you have been there when I needed you, asked me the right questions and left me alone when necessary. Thank you L, R and L.

Table of Contents

Content		Page No
Title Page		
Abstract		**I**
Acknowledgments		**II**
Table of Contents		**IV**
Table of Appendices		**VI**
List of Tables		**VII**
List of Figures		**IX**
Chapter 1	Thesis Introduction	**1**
Chapter 2	Systematic Literature Review: A Review of How Violent and Sexually Violent Men are Most Effectively Managed on Release from Custody	**13**
	Introduction	**14**
	Aims of the Current Review	**16**
	Scoping Review	**16**
	Systematic Literature Review Method	**19**
	Search Results	**26**
	Quality Assessment	**27**
	Results	**30**
	Discussion	**52**
	Conclusions and Recommendations for Practice	**60**

Chapter 3 Empirical Research: Are Community Intervention Services Effective at Keeping High-Risk, Complex, Violent, Male Offenders in the Community for Longer? **64**

Introduction **65**

Methodology **74**

Results **86**

Discussion **100**

Conclusions and Recommendations for Further Research **113**

Chapter 4 Critique of the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) **117**

Introduction **118**

Overview of the CORE-OM **120**

Characteristics of the CORE-OM **122**

Application of the CORE-OM in Forensic Settings **134**

Conclusions **136**

Chapter 5 General Discussion **140**

References **152**

Appendices **173**

Table of Appendices

Appendix	Title	Page No
CHAPTER 2		
Appendix A	Scoping Review Search Terms	**173**
Appendix B	Systematic Literature Review Search Terms by Database	**175**
Appendix C	Additional Articles for Inclusion	**178**
Appendix D	List of Professionals Contacted and Example Contact Email	**179**
Appendix E	Combined Results of CASP Quality Assessment Checklist	**180**
Appendix F	Example Cohort Design CASP Form	**182**
CHAPTER 3		
Appendix G	Memorandum of Understanding	**190**
Appendix H	Variables Requested from North of England Commissioning Support Unit (NECSU)	**198**
Appendix I	Additional Explanation of the Propensity Score Method Used	**199**
Appendix J	Balance Statistics Before and After Matching	**206**
Appendix K	Remaining Kaplan-Meier Survival Curves	**209**
Appendix L	Explanation of the Proportional Hazard Assumption Including Covariates Assessed for Time Dependence	**213**
CHAPTER 4		
Appendix M	CORE-OM Dimensional Breakdown	**215**

List of Tables

Table No	Title	Page No
CHAPTER 2		
Table 1	PICO Framework and Inclusion/Exclusion Criteria	**23**
Table 2	Summary of Included Studies	**35**
CHAPTER 3		
Table 3	Sample Characteristics	**83**
Table 4	Return to Custody and Survival Data by Group Across Each Variable	**87**
Table 5	Recall, Reconviction and Log Rank Test Results	**88**
Table 6	Influence of Individual Covariates on Survival in the Community	**94**
Table 7	Final Regression Model for IIRMS Population	**97**
Table 8	Final Regression Model for IRiS Population	**98**
Table 9	Final Regression Model for Intervention Population	**99**
CHAPTER 4		
Table 10	Coefficient Alpha Denoting Internal Consistency for Non-Clinical and Clinical Samples	**125**
Table 11	Test-Retest Stability in Non-Clinical Samples	**127**
Table 12	Correlations with Referential Measures in Clinical Samples	**130**
Table 13	Means and Standard Deviations for Clinical and Non-Clinical Samples	**131**
APPENDICES		
Table A1	Cochrane Library Initial Scoping Review Search Terms	**173**
Table A2	Cochrane Library Additional Scoping Review Search Terms	**173**
Table A3	Campbell Library Initial Scoping Review Search Terms	**173**
Table A4	Campbell Library Additional Scoping Review Search Terms	**174**

Table B1	Search Terms Used for Systematic Literature Review by Database	**175**
Table E1	Cohort Design CASP Checklist Results	**180**
Table E2	Randomised Control Trial Design CASP Checklist Results	**181**
Table H1	Variables Requested from North of England Commissioning Support Unit (NECSU)	**198**
Table I1	Matching Criteria Used in Relevant Literature, Including the Current Study	**200**
Table I2	Final Matching Methods and Sample Sizes	**202**
Table J1	Balance Statistics Before and After Matching for IIRMS Population	**206**
Table J2	Balance Statistics Before and After Matching for IRiS Population	**207**
Table J3	Balance Statistics Before and After Matching for Intervention Population	**208**
Table L1	Covariates Assessed for Time Dependence or Independence by Group	**213**
Table M1	Dimensional Breakdown of the CORE Outcome Measure	**215**

List of Figures

Figure No	Title	Page No
CHAPTER 2		
Figure 1	Search Results and Screening Process	**29**
CHAPTER 3		
Figure 2	Typical IIRMS Process	**69**
Figure 3	Typical IRiS Process	**71**
Figure 4	Survival to Return to Custody for IIRMS Group	**89**
Figure 5	Survival to Return to Custody for IRiS Group	**90**
Figure 6	Survival to Return to Custody for Intervention Group	**90**
Figure 7	Survival to Reconviction for IIRMS Group	**91**
Figure 8	Survival to Reconviction for Intervention Group	**91**
APPENDICES		
Figure I1	Love Plot of Covariate Balance for IIRMS Population	**203**
Figure I2	Love Plot of Covariate Balance for IRiS Population	**204**
Figure I3	Love Plot of Covariate Balance for Intervention Population	**205**
Figure K1	Survival to Recall for IIRMS Group	**209**
Figure K2	Survival to 100 Days for IIRMS Group	**210**
Figure K3	Survival to Recall for IRiS Group	**210**
Figure K4	Survival to Reconviction for IRiS Group	**211**
Figure K5	Survival to 100 Days for IRiS Group	**211**
Figure K6	Survival to Recall for Intervention Group	**212**
Figure K7	Survival to 100 Days for Intervention Group	**212**

CHAPTER 1

Thesis Introduction

The Current Community Risk Management Problem

The long-term trend in western societies has been towards mass-incarceration (Sturge, 2018; The Sentencing Project, 2019), although this now appears to have peaked, with imprisonment numbers falling in the UK since 2011 (Sturge, 2022) and in the US since 2009 (Carson, 2020). Whilst this change is generally welcomed (Kirk & Wakefield, 2018), the downward trend in worldwide incarceration numbers (World Prison Brief, 2020) significantly increases the burden placed on Parole and Probation services to effectively manage higher numbers of both community sentenced offenders and offenders released from long-term custodial sentences. At the end of 2021, the number of offenders managed by the Probation Service in the UK was 238,500 (Ministry of Justice, 2022), with 862,100 released offenders being managed by Parole services in the US at the end of 2020 (US Department of Justice, 2021). With more people being given community orders or suspended sentences for less serious offences, the logical conclusion is that offenders being released from custodial sentences are convicted of more serious offences.

Once released, recalls to custody and proven reoffending among former prisoners are significant problems worldwide, resulting in serious consequences for the individual, their families, and their wider communities (Adekeye & Emmanuel, 2018; Rakis, 2005). International research illustrates this problem with Petersilia (2011) estimating that of all adult arrests in the US, between 15 – 20% are released offenders and the Social Exclusion Unit (2002) establishing that, of all new recorded crimes in the UK, 18% are committed by released offenders. Furthermore, over 67% of prisoners released across 30 US states in 2005 were re-arrested within three years, and over 76% were re-arrested within five years (Durose et al., 2014). In addition to the risk of being reincarcerated for further offences, released offenders under community supervision, who are treated in much the same way as probationers in the US (Phelps, 2020), must adhere to a strict range of conditions or risk a return to prison for a

technical breach (Doherty, 2016). Antonio and Crossett (2017) report that up to 80% of parole violations are for technical, as opposed to criminal, breaches. With Polaschek et al. (2018) reporting that half of New Zealand's high-risk released offenders are returned to custody within 12 months, and, of these, half do not last 100 days before recall (Nadesu, 2007), the evidence suggests that there is a significant problem when attempting the long-term release of high-risk, complex offenders into the community.

Offenders frequently enter prison with significant difficulties with finances, employment, mental health, and substance misuse (Berghuis, 2018; Kirk & Wakefield, 2018), often leaving prison with the same issues, in addition to the loss of housing, social isolation, and relationship breakdown (Maguire, 2007). This makes reintegration into the community extremely challenging, particularly for long-term prisoners, such as those considered high-risk of general recidivism (e.g., Offender Group Reconviction Scale; Howard et al., 2009; Level of Service Inventory-Revised, Andrews & Bonta, 2003) or potential severity of harm (e.g., Risk of Serious Harm scale, HMPPS, 2022).

Mental Disorders Among Incarcerated Offenders

Mental disorders are more prevalent in offending populations than in the general population (Gottfried & Christopher, 2017) and have also been found – if left untreated – to be linked with higher rates of recidivism (Bonta et al., 2014; Collison & Lynam, 2021). A recent report into prison mental health services in England states that nine out of ten prisoners are incarcerated with at least one mental health or substance misuse problem (Duncan, 2023). The same report confirms that only a fraction of these prisoners receives the support of mental health services in custody and that personality disorder was the most common diagnosis after anxiety and/or depression and psychosis (Brader, 2023). This suggests that a significant number of prisoners

with complex needs are being released from custody having received little or no treatment or support.

Personality Disorder and Recidivism

Personality disorder is a recognised mental disorder as classified by the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition Text Revision (DSM-5-TR; American Psychiatric Association, 2022) and the International Classification of Diseases, 11th Edition (ICD-11; World Health Organisation, 2018). Personality disorder can be viewed as a relational difficulty, both in the context of relationships with other people and the relationship people have with themselves. According to the DSM-5-TR, "personality disorders are long-term patterns of behavior and inner experiences that differ significantly from what is expected. They affect at least two of these areas: way of thinking about oneself and others; way of responding emotionally; way of relating to other people; way of controlling one's behavior." (APA, 2022, para. 3). Whilst most people with personality disorder do not come into contact with the criminal justice system (West, 2014), Coid (1998) established that borderline, anti-social, narcissistic, and paranoid personality disorders are among the most prevalent in the prison population. It is estimated that the majority of people within the UK prison and probation population experience personality disorder (Ministry of Justice, 2011), with most of these experiencing traits of anti-social personality disorder (ASPD; Singleton et al., 1998). However, the cyclical nature of this diagnosis is acknowledged, with the behaviour that results in people being taken into custody typically being used to diagnose the most prevalent personality disorder in custody, ASPD (Ryan et al, 2022). Furthermore, links have been established between offenders with personality disorder and higher levels of recidivism on release from custody (Coid et al., 2006; Fridell et al., 2008; Hernandez-Avila et al., 2000). It has also been established that comorbidity between personality disorder and substance misuse is a strong

predictor of future offending (Fridell et al., 2008; Walter et al., 2011) and that higher numbers of adverse childhood experiences (ACEs) are associated with higher rates of violence in adulthood (Pournaghash-Tehrani & Feizabadi, 2009) and recidivism (Manchak et al., 2008).

Both in custody and the community, most of the forensic population with traits of personality disorder are unlikely to have a formal mental health assessment (Ryan et al., 2022), meaning that this group are typically not supported by mainstream mental health services. Webster and Gardner (2021) highlight further barriers to engagement with community treatment services, such as released offenders and mental health staff being unwilling to engage in mandated treatment and instability with employment, housing and family relationships preventing meaningful engagement by service users. Given the trauma backgrounds of people with traits of personality disorder, it is also acknowledged that relationships between service users and professionals are fragile, and that high levels of fear and shame can prevent people from accessing community-based services (Ryan et al, 2019). This evidence highlights the increased risk of harm and recidivism among people with traits of personality disorder, whilst also demonstrating the difficulties that community treatment and risk management services experience when trying to engage and support this complex group of released offenders.

As well as recognising the problems associated with mental health, the well-being of released offenders, and the issue of recidivism, there is also an economic consideration. The average annual cost per prison place in England and Wales was £43,213 in 2018/19 (Ministry of Justice, 2019) compared to the cost of supervising released offenders under licence, which was recorded as just £2,380 annually (Ministry of Justice, 2012). From a simple cost perspective, this highlights the need to avoid individuals going back into prison. Estimates by the National Audit Office in 2010 place the cost to society of reoffending by released offenders at between £9.5bn and £13bn (Bruce et al., 2017). The significant difference in cost between incarceration and community supervision, combined with high reoffending and recall rates among complex,

high-risk offenders, suggests that attention needs to focus on community management processes.

The Current Thesis

The research outlined above highlights the scale of the challenge for community risk management services, such as the Probation Service in England and Wales. This challenge is most acute for a sub-group of offenders convicted of violence, assessed as high-risk and with complex emotional needs. Not only is this group challenging to engage and support in the community, often after release from lengthy sentences, but they are also at the greatest risk of causing most harm should they reoffend (Howard et al., 2008). To fully explore how to address the current risk management problem, this thesis focuses on how high-risk, complex, male, violent offenders are managed in the community following release from custodial sentences.

Whilst personality disorder is a central feature of the current thesis, it is acknowledged that this is a difficult topic to isolate, both in the existing literature and when conducting original research. Although personality disorder is slowly breaking free from the shackles of being viewed as untreatable (NIMHE, 2003), there remain significant challenges in accurately diagnosing personality disorders in a forensic population (HMPPS, 2020). These difficulties include limited professionals trained to accurately assess personality disorder (Minoudis et al., 2012), complexity in assessment, particularly when comorbid with substance misuse, and lingering attitudes towards this group of individuals as 'untreatable' or 'difficult' (HMPPS, 2020). This makes the study of this group particularly complex, as they are difficult to identify and there is significant heterogeneity within the personality disorder population (Cavelti et al., 2021). As a result of these difficulties, the current thesis is inclusive of all studies and individuals who can be considered complex, challenging, and high-risk. It is recognised that this is a broad study group, but it is also acknowledged that this is most likely to include the

majority of both diagnosed and undiagnosed personality disorder cases in the forensic population. Each chapter of the thesis is as inclusive as possible of this group of released offenders, whilst being robust in their exclusion of less serious, lower risk, and less complex offenders.

The target demographic in the systematic literature review (SLR) in Chapter 2 is high-risk, violent, and sexually violent offenders considered complex or challenging. The SLR does not use the term 'personality disorder' in its search criteria, as this would likely exclude a significant proportion of the literature on offenders with an undiagnosed personality disorder and would also exclude younger adults. Some studies refer to diagnostic tools such as DSM-V-TR and ICD-11 (or older versions) and others highlight emotional or behavioural complexities, aligned with personality difficulties. This is considered an inclusive approach to examining the literature on the target demographic for this thesis. Further information on search terms and inclusion and exclusion criteria are presented in Chapter 2.

The empirical study, in Chapter 3, focuses exclusively on men screened into the Offender Personality Disorder Pathway (OPDP; HMPPS & NHS, 2020). The screening tool is a non-diagnostic process which identifies offenders with additional complex needs, closely aligned with personality disorder traits (Mawby et al., 2020). In addition to the personality-based items required to screen into the OPDP service, cases are also required to be convicted of a violence offence, according to Home Office Crime Codes (Home Office, 2018) and assessed as high- or very high-risk, according to the Risk of Serious Harm scale (HMPPS, 2022). This ensures that the empirical study focuses exclusively on men convicted of violent or sexually violent offences, considered high-risk of harm with additional complex emotional and / or behavioural needs.

The critical evaluation of the CORE-OM, in Chapter 4, initially examines the tool with reference to a general population and then in the context of primary care patients. After this

more generic examination, the CORE-OM is evaluated for use with both forensic populations and those with a diagnosis of personality disorder. Whilst the CORE-OM is not included as an outcome measure in the empirical research, for reasons explained below, this chapter supports the inclusion of the CORE-OM as an outcome measure for the Intensive Intervention and Risk Management Service (IIRMS) examined in Chapter 3.

The evaluation of risk in violent offenders has developed significantly over recent decades and is concerned with predicting the likelihood of a further offence from a specific individual or group of individuals (Campbell et al., 2009). The current thesis uses risk assessment scores as a means of both identifying relevant literature (Chapter 2) and creating comparable control groups (Chapter 3). Although the CORE-OM, examined in Chapter 4, is not considered a risk assessment per se, it does contain a risk domain, which has been found to correlate highly with clinicians' views of risk in relation to borderline personality disorder patients (Whewell & Bonanno, 2000). It is recognised that the use of risk assessment tools, particularly across different countries, can result in difficulties with comparability. For example, the term 'high-risk' is defined and measured differently across jurisdictions. Some measures focus on the risk of general recidivism, regardless of severity, for example, the Offender Group Reconviction Scale (OGRS; Howard et al., 2009) and the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 2003). Other measures focus on the potential for serious harm to be caused, for example, the Risk of Serious Harm scale (RoSH; HM Prison and Probation Service, 2022). Care is taken in the SLR (Chapter 2) to only include cases at the higher end of the risk assessment scale being used, thus targeting the population of interest to this thesis. As the empirical research in Chapter 3 aims to create experimental and control groups that are as closely matched as possible, both the Risk of Serious Harm scale (HMPPS, 2022) and the Offender Group Reconviction Scale (Howard et al., 2009) are used to match for potential severity of harm and likelihood of recidivism, respectively.

Whilst it is acknowledged that it cannot be guaranteed that the cohort examined in Chapter 2 is identical to the cohort examined in Chapter 3, every effort has been made to maximise their similarity and to emphasis the findings which most closely approximate the population of interest – male offenders convicted of violent or sexually violent offences, considered high-risk and complex with regard to emotional or behavioural indicators.

The scope of the research includes western democratic nations, as they tend to have broadly comparable prison and community management systems. It is recognised that this introduces an element of bias into the thesis and care has been taken to limit any conclusions drawn to those regions that the literature is drawn from, and the research was conducted in.

The aims of the thesis are threefold:

- To present a worldwide understanding of how high-risk, complex, male, violent offenders have been managed in the community over recent years and which approaches appear to be most successful in reducing reoffending and prolonging release.
- To explore current intervention and risk management services across the UK for high-risk, complex, male, violent offenders to fully understand their efficacy and to consider if particular sub-groups of this population respond better or worse than other groups.
- To consider the most useful and accurate measures of efficacy in risk management and intervention services for high-risk, complex, male, violent offenders and specifically to evaluate the use of the CORE-OM in measuring progress with this group.

Brief Chapter Summary

Chapter 2

A systematic literature review is presented in Chapter 2. Using strict criteria to identify published studies from around the world, the literature review examines the efficacy of treatment, support, and risk management services in the community for high-risk, complex, violent men released from custody. Although *complexity* is difficult to isolate and measure, attempts are made to focus on groups of individuals considered challenging from a personality, emotional or behavioural perspective. Such difficulties overlap significantly with the construct of personality disorder. The aims of the review are to identify which reentry programmes are effective in reducing undesirable outcomes, which interventions appear not to be effective, and whether specific approaches appear to be more efficacious for sub-groups of this population. The review supports the overall aims of the thesis by providing a comprehensive analysis of what appears to work with this population in terms of both reducing recidivism and prolonging release. Furthermore, the review informs both the methodology and choice of outcome measures in the design of the empirical study in Chapter 3.

Chapter 3

Chapter 3 presents the first large-scale, multi-site, evaluation of UK community risk management services for high-risk, complex, violent men released from custody. This original study uses data from six established community intervention services that support violent and sexually violent men with traits of personality disorder on release from custody. The design of this study is directly informed by the systematic literature review in Chapter 2 and aims to focus on the same population of offenders. However, for the reasons described above and explored below, it is challenging to compare groups of men with personality disorder or who

are regarded as complex both emotionally and behaviourally. The methodology reduces imbalance between treatment and control groups in order to allow for an accurate comparison of outcomes. The study uses a range of outcome measures and employs survival analysis to go beyond the more typical binary outcome analysis relating to *success* or *failure*. The study contributes to the overall aims of the thesis by drawing conclusions about the efficacy of existing intervention and risk management services in the UK, establishing for whom they work best and worst, and offering suggestions for future research in this area.

Chapter 4

Chapter 4 evaluates the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM; Evans et al., 2000). The CORE-OM is a 34-item self-report questionnaire identifying symptoms indicative of global distress. Whilst it was initially validated for use with a general population (Connell et al., 2007) and with primary care patients (Evans et al., 2003), it is increasingly being used with both secure and community-based forensic populations (Kotterbova & Lad, 2022; Perry et al., 2013). The CORE-OM is routinely used as an outcome measure in UK community-based risk management services (HMPPS & NHS, 2021) and is evaluated in the current thesis as a means of better understanding the efficacy of such services. Whilst outcome measure data, including CORE-OM scores, were requested for inclusion in the empirical study, the quality and completeness of data was not sufficient to include in the research. Despite this, the critique of the CORE-OM in Chapter 4 contributes to the overall aims of the thesis by adding to our understanding of robust and accurate outcome measures with which to evaluate services for this complex group of offenders. It also addresses concerns raised in Chapter 3 about the limitations of measuring intervention services using recidivism measures alone. The critique of the CORE-OM sits towards the end of the thesis as, by this stage, attention is focused on the future of service evaluation. As the CORE-OM has now

become a formal outcome measure for the services evaluated in Chapter 3, it was felt that the end of the thesis was an appropriate place for discussions about how the CORE-OM can contribute to our understanding of the efficacy of interventions and how they can be adapted to meet the needs of clients.

Chapter 5

Finally, Chapter 5 provides a general discussion of the key themes and findings from the literature review, original research, and psychometric critique, with reference to the aims of the thesis. An examination of the current IIRMS model is presented, with suggestions being made for prioritising specific groups of service users for IIRMS. In addition, emphasis is placed on existing aspects of the IIRMS model in accordance with findings from this thesis, particularly regarding the type of interventions offered and the measures used to assess efficacy. Harm reduction is also discussed, overall conclusions are drawn, and areas for further research are identified.

CHAPTER 2

Systematic Literature Review

A Review of How Violent and Sexually Violent Men are Most Effectively Managed on Release from Custody

Introduction

Re-entry, the term referring to a return to the community from a custodial sentence, has attracted increasing policy attention over recent decades (Bouffard & Bergeron, 2006), leading to significant financial investment, particularly in the US (Travis, 2005). However, criminal justice agencies have failed to develop a singular, evidence-based method of successfully releasing offenders into the community (Braga et al., 2009). Various theories have been drawn upon and approaches have been adopted, including deterrence theory (Polaschek et al, 2018), the risk-need-responsivity (RNR) model (Bonta & Andrews, 2016) and a resettlement approach (Polaschek & Yesberg, 2018). Whilst the rehabilitative approach offered by the RNR model has gathered momentum within criminal justice, all too often services have been delivered as single interventions aimed at a specific area of need (Lattimore & Visher, 2009). These have included skills acquisition to assist with employment, substance misuse support, assistance with housing and social or emotional support. As Lattimore and Visher (2009) highlight, "the complexity of the disadvantages confronting prisoners after release means that individual offenders often require more than a single programme of intervention" (p.5). This is summarised by Travis and Petersilia's (2001) earlier suggestion that the consequences of long-term rises in imprisonment rates are "a decline in preparation for the return to community" (p.300).

The group of men focused on in the current review are more complex than the general prison population, as they have been classified as high- or very high-risk, according to risk assessment procedures in the local area. Inclusion and exclusion criteria also ensure that all, or the majority, of the populations under review have been convicted of a violent or sexually violent offence. In a systematic analysis of 62 surveys across 12 countries, Fazel and Danesh (2002) established that 65% of male prisoners surveyed met the diagnostic threshold for any personality disorder, and 47% of those surveyed met the criteria for anti-social personality

disorder. Focusing on those offenders considered a high-risk of both reoffending and harm, Bell et al. (2006; cited in Howard et al., 2008) established that over half of UK male prisoners meeting the criteria for Dangerous and Severe Personality Disorder (DSPD) displayed the combination of anti-social and borderline personality disorder traits, with only 10% of those not meeting the DSPD criteria displaying this combination of traits. Howard et al.'s (2008) study supports the notion that men meeting the DSM-V criteria for both anti-social and borderline personality disorders appear to be more strongly associated with criminality and with violence in particular. This indicates that a significant proportion of cohorts under review in this paper will experience personality difficulties, a group widely recognised as challenging for professionals to manage (e.g., Duggan, 2009). However, it is acknowledged that there are difficulties in ensuring that each study includes a comparable sample, due to different descriptions of risk levels, offence types and personality functioning. These caveats are fully explored below.

Re-entry is particularly problematic for individuals with personality difficulties as reintegration into society requires interaction with others and personality disorder is best understood as a relational and attachment disorder (HMPPS & NHS, 2020). Furthermore, it is reported that those with a primary diagnosis of personality disorder are often excluded from health services, including some medium secure specialist forensic units (National Institute for Mental Health in England, 2003). Despite significant efforts to improve access to mental health services for this group, it remains the case that accessing secondary care services in the UK for personality disordered offenders continues to be challenging (Craissati et al., 2021).

Whilst there is no confirmed causal link between adverse childhood experience (ACEs) and personality disorder diagnoses, the emerging literature suggests a strong relationship between ACEs and subsequent juvenile criminality associated with ASPD (Craig et al., 2017; DeLisi et al., 2019). Fox et al. (2015), examining over 22,000 US juvenile offenders, established that

for each additional ACE a child was exposed to, they became 35 times more likely than the group with fewer ACEs to be considered a 'serious, violent and chronic' juvenile offender. This suggests that the cohort being examined in the current review will have been exposed to significantly higher levels of childhood trauma than the general population. The complexity this adds to the re-entry and resettlement process is summarised by Stensrud et al. (2019), who "offer a caution to re-entry programs that place people in housing situations with minimal consideration given to trauma-informed planning" (p.202). Perhaps the complexity and diversity among this population of high-risk, violent offenders who are highly likely to have experienced early life trauma and have developed traits associated with personality disorder are the reasons it has been so challenging to establish a single approach to successful reintegration into the community.

Aims of the Current Review

The aim of the current review is to systematically examine the global literature on the treatment, support, and management of high-risk, male, violent and sexually-violent offenders released into the community following a prison sentence. Specific aims are:

- To establish whether existing re-entry processes have been successful in reducing recidivism.
- To establish whether particular approaches to re-entry appear to be more successful for specific groups of offenders than others.

Scoping Review

To justify the current review, a range of searches were conducted across electronic databases to establish whether any previous reviews existed in this area. Initially, Cochrane and

Campbell Library searches were conducted on 14th October 2019 and were updated on 9th January 2023, with both limited to reviews only. Search criteria for this scoping review can be seen in Appendix A. Cochrane Library searches returned 10 reviews for consideration. Having examined these in detail, none were deemed relevant or met inclusion criteria for the current review (see Table 1).

Eight of these reviews were excluded for the following reasons: referred to specific substance misuse treatment programmes (4), involved an exclusively female cohort (2) or involved a youth or adolescent cohort (2). One of the remaining reviews (Dennis et al., 2012) focused on re-entry for men convicted of a sexual offence, although further examination revealed that four of the ten included studies involved custodial or secure hospital treatment and only five reported a primary outcome measure related to recidivism. Although this review was excluded on this basis, it is worth noting that the authors found 'promising' results in two of the studies, with other studies either not recording recidivism or finding no difference between intervention and control groups. The clarion call from this study was for further randomized control trials of this population re-entering the community from custody. It should also be noted that, despite being published in 2012, the most recent study included in this review was from 1997, suggesting that a robust update is due. The final review (Gibbon et al., 2020) focused on 19 studies assessing the efficacy of interventions for people diagnosed with anti-social personality disorder. This review was excluded as half the included studies were for substance misusing participants, over half the studies included female participants and a third of studies examined interventions that occurred in secure settings (hospital or prison).

The Campbell Library search returned two reviews, neither of which met the inclusion criteria.

Following a search of the general literature, described below, a further search was conducted to ensure that no relevant reviews were missed. 'Re-entry' was not included in the initial search terms, although this established itself as a common term for the process under examination. A

further search of both the Cochrane and Campbell Library's was conducted on 19th December 2019 simply using the terms 'reentry' or 're-entry'. This was updated on 9th January 2023. The Cochrane Library search returned four reviews, none of which were relevant, as they referred to specific medical procedures. The Campbell library search returned no reviews.

Although the search of the general literature excluded reviews, a number of meta-analyses and systematic literature reviews were returned and considered for inclusion in the scoping review. Further reviews referenced in the general literature were also considered for inclusion in the scoping review. The majority of these reviews related to correctional or custodial treatment (e.g., Aos, Miller & Drake, 2006; Lipsey & Cullen, 2007; Mackenzie, 2006) and were excluded on this basis. However, two reviews were of particular interest as they initially appeared to meet the criteria for inclusion (Berghuis, 2018; Seiter & Kadela, 2003).

Berghuis (2018) examined the effectiveness of re-entry programmes designed to reduce recidivism among adult male offenders. Although a very relevant and useful study, there are a number of ways in which the current review differs. Berghuis' inclusion criteria stipulated that all studies must be randomised control trials (RCTs). Whilst the focus on RCTs ensured methodological robustness, it is also known that designing this type of research with this client group is challenging and can result in delays and financial implications (Nathan et al., 2019). As well as being challenging to design, there are also ethical and legal concerns around the random assignment of high-risk, violent offenders to treatment or control groups (Procter, 1996). The current review, whilst not excluding RCTs, is open to a wider range of studies which includes the most common form of observational cohort design when studying this client group. Berghuis' initial literature searches were conducted in May 2014, with an updated search in March 2016, indicating that there are almost seven additional years of published research included in the current review. Although Berghuis studied adult male offenders released from prison, the current review focuses on the most complex group of individuals that

are released from prisons worldwide; high-risk men convicted of violent or sexually-violent offences. As described above, this cohort are significantly more likely to have experienced trauma in their early years which may have contributed to the development of traits associated with personality disorder, making their re-entry journeys more complex for both themselves and the professionals supporting them.

Seiter and Kadela (2003) searched the published and unpublished literature between 1975 and 2001 and examined the efficacy of 32 re-entry programmes targeting male offenders across North America. They report reductions in recidivism associated with vocational and work-release programmes, halfway houses and some drug treatment programmes. Only five of the studies reviewed focused exclusively on violent or sexually-violent offenders and only one of these established a small reduction in recidivism among the treatment group. Berghuis (2018) is critical of the review due to poor matching of groups, selection bias due to high attrition rates and inexact reporting of results. The review also examines literature from between 22 and 48 years ago and does not look specifically at high-risk, male offenders. The current review focuses on a narrower cohort and brings a review of the literature up to date.

Systematic Literature Review Method

Search Strategy

Following the scoping review, a search of the general literature was conducted using a range of electronic databases specialising in literature on psychology, sociology, medicine, health science, social science and humanities. These searches were conducted on 24th October 2019 and were updated on 12th January 2023. The databases used were Ovid PsycINFO (1967 to January week 2 2023), Ovid Embase (1974 to 2023 January 12), Ovid MEDLINE® (1946 to January week 2 2023), Web of Science (Core Collection, 1900 to 2023), SCOPUS (1970 to

current), National Criminal Justice Reference Service Abstracts (NCJRS; 1975 to current), Applied Social Sciences Index and Abstracts (ASSIA; 1987 to current) and Social Science Database. Journal articles were the focus of searches and books, technical reports, websites, and podcasts were all excluded, as they are not the formats in which robust, detailed, research on this population is typically published. The search strategy aimed to return all published studies which specifically compared post-release intensive/enhanced supervision and/or treatment for high-risk violent and sexually violent men with a similar group managed via 'Probation as usual'. Specific search strategies varied slightly due to differences in database structure and search capacity. Search terms for each database can be viewed in Appendix B, although most databases were searched using the following search terms:

Keyword Search: ((Crim*) OR (offend*)) **AND** ((release*) OR (parole*) OR (probation*)) **AND** ((manag*) OR (supervis*) OR (rehabilita*) OR (treat*) OR (interven*) OR (diver*)) **AND** (("what works") OR (effectiveness) OR (efficacy) OR (outcome) OR (evaluat*)). *Limits:* Adults only / journals only / males only / English language.

A Note on the Exclusion of Personality Disorder as a Search Term

Although the overarching aim of the current review is to examine the efficacy of risk management processes for some of the most complex men involved in the criminal justice system, many of whom will meet the diagnostic criteria for personality disorder (PD), the term 'personality disorder' was not used to search the literature or stipulated in the inclusion criteria. The main reason for this decision was that PD diagnosis remains a complex process which continues to be inaccessible to non-clinical staff in prison and secure-hospital settings (Minoudis et al., 2012), indicating that it most often goes undiagnosed. Trial search strategies indicated that including personality disorder (and associated variants) significantly reduced the range of literature returned, thus reducing the scope of the current review. It is noteworthy that

the review includes literature prior to 2003, when the National Institute for Mental Health in England released *Personality Disorder: No longer a diagnosis of exclusion.* Prior to this paper, and sometimes more recently, offenders with a diagnosis of PD were likely to have been excluded from re-entry programmes. By avoiding this term in the search criteria, it is hoped that the literature returns a broader cross-section of studies to include men considered complex both emotionally and behaviourally. Furthermore, guidance from DSM-V-TR (APA, 2022) advises that PD should be diagnosed with caution prior to the age of 25, suggesting that using the term personality disorder in the current search criteria would risk excluding literature on younger adult cohorts.

Additional Literature Sources

It is acknowledged that, despite being as comprehensive as possible, a single search strategy will not return all relevant literature within a specified area. Examination of the most relevant papers returned by the literature search revealed eight further papers of interest (Appendix C). These were all obtained and added to the literature database for screening. An open-source search was also conducted on Google Scholar guided by the search strategy described above. No additional papers were added to the literature database from this search.

In addition, a number of experts in the field of treatment, intervention and management of high-risk violent offenders were contacted. These experts were identified by a combination of the primary researcher, who has over 20 years' experience working in offender risk management, and suggestions from experienced colleagues, who are all clinical leads of community intervention services. Details of these individuals and the contact method used can be viewed in Appendix D. Several suggestions for papers to review and areas to explore were made but this did not identify any relevant studies in addition to the general literature search.

Inclusion and Exclusion Criteria

Using the comprehensive database of literature identified, studies were initially screened for relevance. Inclusion and exclusion criteria were used to guide this process in order that only the most relevant studies remained. Table 1 summarises the inclusion/exclusion criteria using a Population/Intervention/Comparator/Outcome (PICO) framework.

The inclusion and exclusion criteria were designed to ensure only the most relevant research was identified for the quality assessment and review process. The current review focuses on high-risk, adult, male offenders convicted of either a violent or sexually violent offence. The population inclusion criteria reflect this and specifies that the cohort will have been released from a custodial sentence. This ensured the removal of a substantial sub-section of the literature which focuses on supporting men under the supervision of Parole/Probation services as they serve a community or suspended sentence. The exclusion criteria also ensured that all studies focusing on offenders with serious mental illness, predominantly psychotic disorders, were removed from the database. This group of offenders require a different form of specialist, often medical, intervention which is not the focus of typical re-entry and risk reduction interventions in the community. Although the inclusion criteria stipulate male only cohorts, studies involving majority male cohorts were included, as it was felt that a substantial amount of useful data would otherwise be lost. Where a small number of female offenders were included in the cohort, care is taken to acknowledge and account for this when synthesising the results.

The intervention was required to be delivered in the community, although studies that began preparation for re-entry in custody or secure hospitals were included, as it is recognised that the transition process into the community begins in a secure setting. A judgement was made as to whether the intervention was *primarily* delivered in custody, in which case it was

Table 1

PICO Framework and Inclusion/Exclusion Criteria

PICO Framework	***Inclusion Criteria***	***Exclusion Criteria***
Population	Adults	Studies exclusively focusing on youth, adolescents or children.
	Offenders defined as high- or very high-risk of reoffending/harm released from custody	Cohorts of low/medium risk offenders only or those released/discharged from mental health establishments.
	Convicted of a violent or sexually violent offence	Cohorts sentenced to a community order/suspended sentence
	Majority male cohort	Female only or majority female cohorts
		Studies focusing on serious mental illness and those found not-guilty by reason of insanity (or equivalent regional term)
Intervention	Community based	Treatment/intervention delivered in custody, secure hospital or other involuntary environment
	Intervention, re-entry programme or diversionary project	Specific substance misuse reduction or employment programmes
	Structured programme/intervention	Studies exclusively or predominantly involving pharmacological intervention or other external control methods, such as Polygraph testing, GPS tagging or intensive monitoring/supervision
	Managed by the Probation Service (or equivalent), other Public-Sector organisation or not-for-profit 3rd sector organisations.	Managed by private for-profit organisations.
Comparator	Matched Probation as usual group	No comparator/control group
	Matched service users on waiting list	Unmatched/poorly matched control group
Outcome	Recidivism must be an outcome measure (this can include re-arrest, reoffending, recall, technical	Recidivism not used as an outcome measure or poorly measured

	breaches, revocation of licence etc.)	
Other Criteria	Year of publication: 1990 – January 2023 Language of publication: English Document type: peer-reviewed or non-peer-reviewed journals/articles Subjects: humans only (for medical databases	Document type: Books, technical reports, websites, Podcasts

excluded, or whether this was a preparation phase for the intervention to be delivered in the community, in which case it was included. The intervention was required to be structured, in that it was a planned series of contacts with the intervention group and differed in a substantial way to Probation as usual. The intervention also had to be delivered by a Government funded department, most often Parole or Probation services, or third-sector, not-for-profit organisations. This was stipulated to avoid situations in which the objectivity or unbiased nature of research was called into question. Interventions predominantly involving medication or forms of external control, such as Polygraph testing, GPS tracking or other enhanced monitoring techniques, were excluded as they increase the likelihood of detection without having any long-term effect on an individual's risk.

It was important to ensure that the intervention under examination was compared with a matched or similar control group, to ensure that the results of the study could be reasonably attributed to the intervention. Studies with no control group or with poorly-matched control groups were excluded from the study, as there are numerous confounding variables which can lead to apparently positive or negative outcomes with this population. Treatment and control groups were considered poorly matched if no attempt was made to ensure comparability between groups on the basis of key variables, if an RCT design was not used, if groups were considered markedly different in situation, for example released offenders vs. community

sentenced offenders, or if the control group was from the general population. It was also essential that recidivism, defined as a relapse into offending behaviour after punishment for previous offending, was the primary outcome measure in all included studies. Although it is acknowledged that recidivism can be measured in numerous different ways and that reconviction is not always the most accurate proxy for recidivism (Antonio & Crossett, 2017), the main aim of stipulating this as inclusion criteria was to ensure that each study was comparable. The different ways of approximating recidivism are discussed below.

Finally, the year of publication was limited from January 1990 to January 2023, to capture a wide range of studies during the period in which re-entry has risen higher up the political and economic agenda. 1990 also coincided with the development of structured treatment programmes in the custodial estate in the UK (Falshaw et al., 2003), which increased the provision of community programmes over subsequent years. English was stipulated as the publication language for pragmatic reasons, as time and financial constraints meant that translation from other languages was not possible.

It is recognised that it is not possible to guarantee that each included study contained comparable cohorts of released offenders, particularly in relation to risk level, offence type and personality functioning. Not all studies provided clear descriptions of each of these areas, and it is also acknowledged that each region uses different risk assessments and offence categories. To ensure that as much information was available for analysis as possible, an inclusive approach was taken when descriptions were somewhat unclear. To minimise any bias this may cause, these issues are fully acknowledged in the discussion section below.

Search Results

Initial searches in October 2019 returned 1,406 records across all eight databases and a further 222 records were returned from the updated searches in January 2023, resulting in 1,628 potentially eligible records. Duplicates between databases were identified using a combination of DOI numbers and article titles. As DOI numbers were not available in all cases and because some databases returned titles in different formats, not all duplicates were able to be removed with this strategy. More detailed analysis of the results enabled further duplicates to be removed manually. Although several duplicates are likely to have remained in the database, this method ensured that no unique articles were removed. This process led to 673 duplicates being removed, leaving 955 studies eligible for inclusion.

An initial screening process was undertaken during which titles and abstracts were reviewed for relevance to the area under review. To minimise the degree of subjectivity in the screening process, 10% of the sample (95 of 955 studies) were second scored by an experienced researcher, educated to post-graduate level, who is familiar with the current offending and rehabilitation literature and is unconnected to the study. This resulted in inter-rater reliability (IRR) of 89% which is well above the cut-off of 80% generally considered acceptable in systematic literature reviews (Belur et al., 2018). All but one of the disagreements were due to the paper in question being excluded by the primary researcher and included by the second researcher. These were all resolved through discussion which provided more detail about the inclusion and exclusion criteria. Whilst this did not lead to any changes in the PICO criteria, it did result in more clarity, for example, by stating that participants in the study had to have previously been incarcerated. The final disagreement was resolved through further discussion of the PICO criteria. The initial screening process removed a further 867 studies, leaving 88 studies to be examined in further detail.

At this stage, eight further papers were added to the database following contact with relevant professionals, open-source searches (n=2) and examination of reference lists from articles returned by the systematic search (n=6). These papers are listed in Appendix C. These articles were added at this stage because it had already been decided that they were relevant to the area of interest, yet they required a more detailed screening process to determine whether they should be included in the final group of papers for review. This left 96 studies to be re-screened.

The second screening process was undertaken using the PICO inclusion and exclusion criteria to determine whether studies focused on the specific area of interest and provided results that added to our understanding of 'what works' with this client group. This process involved analysis of abstracts and full text articles, leading to 86 papers being removed and 10 studies being eligible for quality assessment.

Quality Assessment

There is a myriad of quality assessment tools available to guide decision-making about the robustness of research in the human sciences. The chosen tool for this review is provided by the Critical Appraisal Skills Programme (CASP; Critical Appraisal Skills Programme, UK, 2018), who have developed a range of checklists to help assess the trustworthiness, relevance, and results of different study designs. Two of their checklists, specifically developed to guide the process of quality assessment with cohort studies and randomised control trials, have been used in the current review. Unlike many other assessment tools, the CASP checklists do not provide an overall 'score', rather they focus the user on critical questions about research allowing them to come to their own judgements about the quality and utility of each paper. Questions are organised within three sections, the first of which focuses on the validity of the research. The aim of this section is to facilitate decisions regarding bias in the research to draw

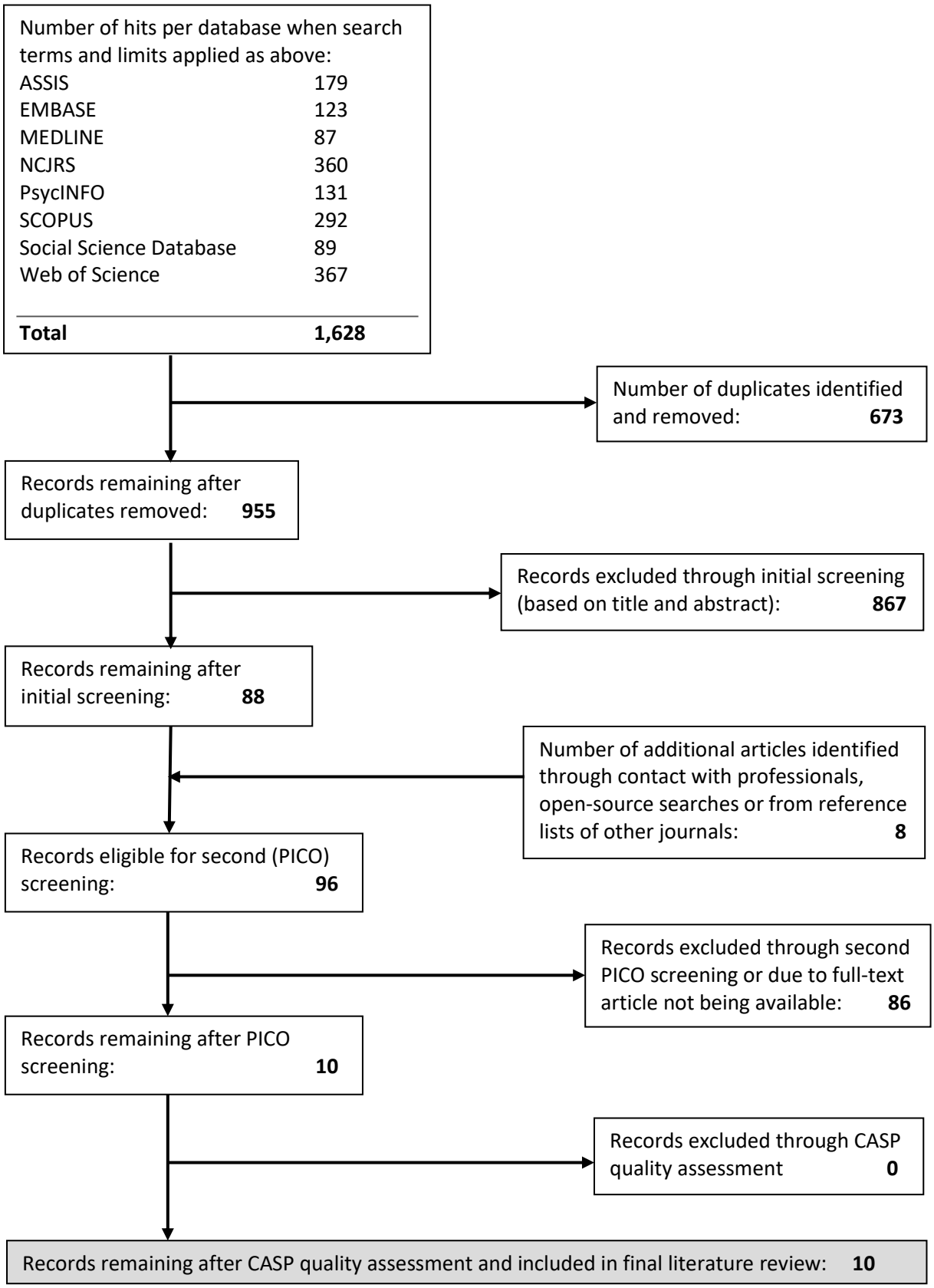
conclusions about methodological quality. The second section of the checklists focuses on the results. The aim here is to guide decisions about the clinical importance of the results and to examine any uncertainty about them. The final section of the checklists focuses on how useful the results are. This allows users to consider the application of each study to the wider research question under review. An example of a completed CASP form can be viewed in Appendix F.

The decision to use the CASP 'guidance' approach to quality assessment, rather than a 'score-based' approach, was made primarily because of the population under investigation. High-risk released offenders are a difficult group to study as they experience high levels of personality dysfunction (Fazel & Danesh, 2002; Singleton et al., 1998), mental health difficulties (Fazel & Seewald, 2012) and can be difficult to engage in follow-up studies. This means that observational designs have been traditionally used to study this group, making it hard to control for confounding variables. In addition, searches resulted in relatively few studies in this area. It was felt that a guidance approach to quality assessment would be more flexible regarding the studies that were included, whilst highlighting any design and methodological concerns raised.

Each of the ten studies which remained eligible following the application of PICO criteria were assessed using either the cohort study or randomised control trial checklist, depending on the study design. A summary table of CASP responses for each study can be viewed in Appendix E. Once the guidance had been followed in full, a rating scale was introduced to allow for comparisons of quality across studies. Each of the guidance criteria was answered '*yes*', '*no*' or *'cannot tell'*. The latter two answers were treated the same, as it meant that the area in question was either not adequately accounted for or was not discussed fully in the paper. As the cohort and RCT checklists used different numbers of questions, percentage scores were calculated for affirmative answers. When making decisions on the quality of each study, attention was paid to the first section of the checklists, as this focuses on methodological flaws. If papers did not meet the criteria in this section, they were removed from the process. No

Figure 1

Search Results and Screening Process (search dates January 1990 – January 2023)



studies were removed at this stage. Studies scoring over 80% were rated *strong*, as they were able to answer positively against all but one or two of the criteria. Studies scoring between 70-79% were rated *moderate*, missing up to three of the criteria. All papers were rated as *strong* or *moderate*, meaning that no studies were removed from the review following the quality assessment process. The final number of studies included in the literature review is ten. Figure 1, above, summarises the process of arriving at the studies included in the final review.

Results

Study Design

Seven of the studies under review adopted a non-randomised observational cohort design. This, most commonly, involved using a retrospective sample of offenders released from custody into community treatment or supervision programmes and comparing outcome measures with a control group who were managed through 'Probation as usual'. Three of the studies adopted a randomised control trial (RCT) design, in which the cohort under investigation were randomly assigned to an intervention group or to a control group. Outcome measures were then compared between the two groups. Six of the reviewed studies were conducted in North America (5 in the USA, 1 in Canada), three in the UK and one in the Netherlands.

Population

To meet the inclusion criteria, the cohort under investigation had to be high-risk violent or sexually violent males (or majority male) released from a custodial sentence. It was not always possible to determine the exact risk level of the sample from the write up and it is also acknowledged that both within and across countries there will have been a range of methods

used to assess for risk level. Where there was some doubt about the exact nature of the cohort under investigation, a judgement was made based on the available information. In such cases, the decision-making process is reported below. An inclusive approach was taken to ensure that relevant literature was not lost and the impact of these decisions on the robustness of the current review are discussed below.

Narrative descriptions of the sample were often provided and have been grouped in the following way. Four studies described their sample as '*highly active violent offenders'* (Braga et al., 2009, p. 421), *'serious and violent offenders'* (Bouffard & Bergeron, 2006, p. 9; Lattimore & Visher, 2009, p. 14) and *'considered to have personality disorders who were identified as high or very high risk of harm to others'* (Nathan et al., 2019, p. 2). The three studies focusing on men convicted of a sexual offence (Hanson et al., 2004; Hedderman & Sugg, 1996; Procter, 1996) all stated that convictions were exclusively or majority for contact offences, indicating that they were considered serious in nature. Risk levels were not reported, and it is acknowledged that these men may not all have been considered high- or very high-risk of harm. However, these studies were included because almost all non-contact offenders were excluded from the cohorts and because it is known that many serious charges for rape result in convictions for indecent assault or sexual assault (Grace et al., 1992). Antonio and Crossett (2017) described their population as *'medium to high-risk'* (p.517), with the remaining two studies being less specific about their sample. Shaul et al. (2016) state that they examined released offenders who had committed at least two previous offences, not including driving convictions. Of this cohort, 52% were convicted of a violent offence and the mean level of self-reported crime in the year prior to arrest was more than 80 offences per person. A decision was made to include this study as it involved prolific, repeat offenders, the majority of whom were convicted of a violent offence. It also excluded some offences which were clearly not of interest to the current review. Clark (2015) focused on offenders previously re-incarcerated

following a failed parole period, due to either the commission of a new offence or a technical breach. Thus, the cohort in this study are high-risk of release violation. Unlike several other large-scale re-entry programmes in the US (e.g., Prisoner Re-entry Initiative, Minnesota Department of Corrections, 2011; Serious Offender Accountability Restoration Project, Minnesota Department of Corrections, 2006), The High-Risk Revocation Reduction (HRRR) programme examined by Clark (2015) was less restrictive regarding eligibility, allowing both men convicted of a sexual offence and men usually managed by intensive supervision processes access to the programme. Although this resulted in a wider range of offences being included, most convictions were for 'person' or sexual offences, justifying its inclusion in the current review.

Three of the studies under review (Hanson et al., 2004; Hedderman & Sugg, 1996; Procter, 1996) focus exclusively on cohorts of men convicted of a sexual offence and are also the three oldest studies included in the review. The remaining seven studies focus largely on men convicted of 'person' or violent offences.

Four studies (Bouffard & Bergeron, 2006; Clark, 2015; Nathan et al., 2019; Procter, 1996) involved relatively small intervention and/or control groups of fewer than 100, suggesting that care needs to be taken when generalising results across larger populations. A further four studies (Braga et al., 2009; Hanson et al., 2004; Hedderman & Sugg, 1996; Shaul et al., 2016) used moderately sized intervention and control groups of between 100 – 500, making generalisability more feasible, once the caveats of each study are considered. Two studies (Antonio & Crossett, 2017; Lattimore & Visher, 2009) had access to large data sets resulting in intervention and control group samples of over 1,000 and 2,000 participants, respectively. Whilst this does not guarantee an ability to generalise results across similar populations, it increases confidence when doing so. Conversely, having total sample sizes running into the

thousands makes it harder to control for confounding variables, such as treatment or support accessed independent of the intervention under review (Hanson et al., 2004).

Follow-up Period

Deciding on a follow-up period for recidivism research has implications for the results (Hoffman & Stone-Meierhoefer, 1980), and appropriate follow-up periods are likely to vary by offence type. As the base rate for sexual offending is relatively low (Hanson & Bussiere, 1998; Hanson & Thornton, 2000) it is anticipated that longer follow-up periods are required to detect any potential treatment effects. As a high proportion of violent offenders are reconvicted in a relatively short space of time (Durose et al., 2014; Langan & Levin, 2002; Nadesu, 2007; Polaschek et al., 2018), shorter follow-up periods are appropriate. Two of the three studies focusing on men convicted of a sexual offence had the longest follow-up periods of 12 years (Hanson et al., 2004) and 5 years (Procter, 1996). It should be noted that not all of Hanson et al.'s (2004) cohort were followed for the full 12 years, although seven-year follow-up data are also reported in the paper. The other paper studying treatment for men convicted of a sexual offence (Hedderman & Sugg, 1996) followed participants for two years, although five and ten-year follow-ups were planned. Five of the remaining studies followed participants for between 12 and 24 months. Two further studies (Antonio & Crossett, 2017; Braga et al., 2009) followed participants for 36 months, a period in which any potential treatment effect is expected to have been detected.

Outcome Measures

There are a range of measures which can be used as proxies for recidivism, but it is accepted that all have their flaws, as many of the type of offences under review go unreported or

unconvicted (Hedderman & Sugg, 1996). Travis and Visher (2005) suggest that re-arrest for a new crime is the most accurate proxy for recidivism, as all other measures are influenced by sentencing policy, court decisions or variations in enforcement processes. That said, arrest rates are not always available, particularly for offences which have very low convictions rates, such as sexual offences. Six of the studies under review used official re-conviction data as the primary outcome measure, although Procter (1996) only looked at re-conviction for a sexual offence. Clark (2015), in addition to re-conviction for any crime, used re-arrest and re-incarceration as proxies for recidivism. Re-arrest data were also used by three studies, with Lattimore & Visher (2009) also using re-incarceration data as an outcome measure.

Antonio & Crossett (2017) focused on re-incarceration, either for a new offence or for a technical breach of parole, as the primary outcome measure.

The range of different outcome measures used, in addition to the different study designs adopted, makes it difficult to compare these studies directly and impossible to pool the results for quantitative analysis. Thus, results are presented individually, and themes are extracted for discussion below.

Table 2

Summary of Included Studies

Authors, Year of Publication and Country	**Title**	**Aims**	**Design**	**Intervention Group**	**Control Group**	**Follow-Up Period**	**Outcome Measures**	**Results**	**CASP Quality Rating**
Antonio & Crossett (2017) USA	Evaluation the effectiveness of the national curriculum and training institute's "Cognitive Life Skills" (CLS) program among parolees supervised by the Pennsylvania Board of Probation and Parole.	To examine the effects of participation in CLS on re-incarceration or revocation of parole compared with a group of offenders who were not offered this intervention. Secondary aims were to assist policy makers both locally and nationally to deliver effective treatment.	Non-randomised observational cohort.	2115 released medium- to high-risk offenders who completed the CLS group programme in the community. 90.4% were male.	2115 released medium- to high-risk offenders matched for key variables with the intervention group. 90.3% were male.	36 months	Recidivism measured by re-incarceration for a new offence and/or for violating a technical condition of parole.	Recidivism rates significantly lower for intervention group. Results were even stronger when comparing high-risk groups. However, significantly more new offences were committed by the high-risk intervention group. Also established that the probability of recidivism was always higher for the control group than for the intervention group.	Strong

Bouffard & Bergeron (2006) USA	Reentry works: The implementation and effectiveness of a Serious and Violent Offender Reentry Initiative (SVORI).	To examine the impact of participation in a high intensity reentry programme (SVORI) on in-programme and post-parole re-arrest rates in comparison with a matched group subject to 'probation as usual'.	Non-randomised observational cohort.	71 serious and violent offenders aged between 18-35, released to a single urban district in upper-mid-west US. 85.9% were male.	106 matched serious and violent offenders, aged between 18-35, released to other districts within the same state. 84.0% were male.	13 months (400 days)	Recidivism measured by drug/alcohol test results, revocation of parole and re-arrest data (post-parole).	Significantly higher drug testing rates found in the intervention group. However, this group provided significantly fewer positive tests. Intervention group were found to have a significantly lower risk of re-arrest than those in the control group. No significant differences between groups regarding parole revocation and survival time.	Strong
Braga, Piehl & Hureau (2009) USA	Controlling violent offenders released to the community: An evaluation of the Boston Reentry Initiative.	To evaluate the effects of the Boston Reentry Initiative (BRI) on the subsequent recidivism of participants in comparison to a matched control group.	Non-randomised observational cohort.	108 'highly active' violent offenders aged between 18-32, selected from the local jail in 2002.	309 matched violent offenders aged between 18-32, selected from the local jail in 2000.	36 months	Recidivism measured by re-arrest rates. Both 'any arrest' and 'violent arrest' were measured.	Re-arrest rates for any offence significantly lower for intervention group across the three-year follow-up period. Re-arrest rates for violent crime significantly lower for intervention group across the three-year follow up period.	Strong

Clark (2015) USA	Making the most of second chances: An evaluation of Minnesota's high-risk revocation reduction reentry program.	To assess whether a re-entry programme targeting high-risk release violators (RVs) leaving state prison significantly reduced recidivism.	Randomised control trial.	162 adult male offenders who had previously failed on parole (reoffence or technical breach) randomly assigned to intervention group.	77 adult male offenders who had previously failed on parole (reoffence or technical breach) randomly assigned to control group.	12 – 24 months	Recidivism measured by supervised release revocation, re-arrest, reconviction and re-incarceration for a new offence.	Intervention group significantly less likely to have their supervised release revoked or to be reconvicted than control group. Smaller proportion of the intervention group were re-arrested or re-incarcerated for a new offence, although these findings were not significant.	Strong
Hanson, Broom & Stephenson (2004) Canada	Evaluating community sex offender treatment programs: A 12-Year follow-up of 724 offenders.	To examine the effect of community treatment on an unselected cohort of men convicted of a sexual offence, compared with an unselected cohort of men convicted of a sexual offence who were not offered community treatment.	Non-randomised observational cohort.	403 men convicted of a sexual offence and treated via the Community Sex Offender Programme (CSOP) in the community.	321 men convicted of a sexual offence and released prior to roll out of CSOP in their area.	7 years and 12 years	Recidivism measured by official Police records of charges or convictions for new sexual, violent or general criminal offences.	No treatment effect found for sexual, violent or general recidivism after an average follow-up of 12.5 years and a fixed follow-up of 7 years. Quality of treatment provided had no effect on recidivism rates.	Moderate

Hedderman & Sugg (1996) UK	Does treating sex offenders reduce reoffending?	To examine reconviction rates two years post-treatment completion to discover whether treated offenders were convicted of fewer or less serious offences than a control sample.	Non-randomised observational cohort.	133 men convicted of a sexual offence and referred for community treatment in one of seven programmes.	191 men convicted of a sexual offence and given probation orders but not referred for treatment.	24 months	Recidivism was measured using official reconviction data.	Intervention group reconvicted for a sexual offence at a lower rate (5%) than the control group (9%). All but one of the sexual reconvictions in the intervention group were judged to be for a similar or less serious offence than the index offence. Total rate of reconviction for any offence for the intervention group (9%) was lower than the predicted reoffending rate (13%). Total rate of reconviction for the control group (29%) was higher than the predicted reoffending rate (23%).	Moderate
Lattimore & Visher (2009) (unpublished but made available by	The multi-site evaluation of SVORI:	To examine the extent to which Serious and Violent Offender	Non-randomised observational cohort.	1,168 participants across 12 adult and 4 juvenile sites (863 adult	1,223 matched participants assigned to 'Probation	24 months	Access to services, substance misuse tests, re-arrest and	Intervention group received significantly more services and accessed more	Moderate

the US Dept. of Justice). USA	Summary and synthesis.	Reentry Initiative (SVORI) participation improved access to appropriate, robust, integrated services and resulted in better outcomes regarding recidivism.		males, 153 adult females and 152 juveniles).	as usual' group (834 adult males, 204 adult females and 185 juveniles).		re-incarceration rates.	programs than control group. No significant differences between groups for re-arrest and re-incarceration rates for adult males	
Nathan, Centifanti, Baker & Hill (2019) UK	A pilot randomised control trial of a programme of psychosocial interventions (Resettle) for high-risk personality disordered offenders.	To test the feasibility of a randomised control trial of a complex intervention for high-risk personality disordered offenders and to gather evidence of potential treatment effects.	Exploratory randomised control trial.	38 released offenders, over 18, likely to have a personality disorder and deemed in need of multi-agency risk management arrangements, therefore high-risk.	34 released offenders randomly assigned to the control group from the same cohort as the intervention group.	24 months	Recidivism measured using officially recorded Police data for reconvictions cautions and reprimands. Self-reported anti-social behaviour was also used as an	Intent-to-treat analysis indicated that the intervention group were reconvicted at a significantly higher rate than the control group. However, they were also significantly more likely to commit no offences during follow-up. Analysis of those actually receiving treatment indicated that the intervention	Moderate

							outcome measure.	group were significantly less likely to be reconvicted. However, the control group were significantly more likely to commit no offences during follow-up.	
Procter (1996) UK	A five-year outcome evaluation of a community-based treatment programme for convicted sexual offenders run by the Probation Service	To examine the effects on sexual recidivism of a community treatment programme offered to men convicted of a sexual offence, compared to a matched control group who were not offered community treatment.	Non-randomised observational cohort.	54 men convicted of a sexual offence who had completed the community-based 'Cherwell Group' programme between 1989 and 1992.	54 men convicted of a sexual offence, who were matched on key variables to the intervention group but who were not offered community treatment.	60 months	Recidivism was measured by official records of reconviction for a further sexual offence.	No statistically significant differences between groups. Survival analysis indicated positive, but non-significant, results for the intervention group after four and five years.	Strong
Shaul, Koeter & Schippers (2016)	Brief Motivation Enhancing Intervention (MEI) to	To examine the effects of a protocolized MEI on	Randomised control trial.	111 male offenders who had committed at least two	109 male offenders randomly assigned to	12 months	Recidivism measured by both self-reported	No significant differences between groups regarding self-reported recidivism,	Moderate

Netherlands	prevent criminal recidivism in substance-abusing offenders under supervision: A randomized trial.	recidivism rates and time to reoffending for substance misusing offenders supervised in the community compared with a matched control group.		offences randomly assigned (using cluster randomisation) to complete MEI in the community.	'supervision as usual' condition in the community.		reoffending and official records of reconviction within the follow-up period.	official records of recidivism or a combination of these measures. No effect of treatment on time to reoffending within the 12-month follow-up period.	

Data Synthesis

The ten studies under review use a range of designs and outcome measures. Data synthesis is organised by these outcome measures, all of which aim to approximate recidivism using different approaches. Results are initially presented according to the four outcome measures used: *revocation of parole/licence, re-arrest, reconviction,* and *re-incarceration*. Some studies also report results for access to community services and specific housing and employment outcomes. These are reported in the final two results sections.

Of the ten studies included in the systematic literature review, six were conducted in North America (5 USA; 1 Canada), 3 in the UK and the remaining study was conducted in the Netherlands. The average sample size was 890, although the range was between 72 and 4,230 participants. The minimum follow-up period was 12 months, and the maximum was 12 years, although the most common follow-up period was 24 months (n=4). Participants were majority adults, although one study included a small cohort of juvenile offenders, and majority male, although three studies included small cohorts of female offenders. Participants were convicted of violent or sexually violent offences or were considered high-risk, as measured by local assessment tools. Three studies focused exclusively on men convicted of a sexual offence. Each study described the personality and risk complexity of their cohorts in different ways, and these are explored further below.

Recidivism: Technical breach of parole

Three studies reported technical breach of parole rates (Antonio & Crossett, 2017; Bouffard & Bergeron, 2006; Clark, 2015), although each study also examined other recidivism measures. It is noteworthy that as well as including technical breaches of parole, Bouffard and Bergeron (2006; CASP rating strong) included in their data those who were re-arrested during parole.

24.3% of the Serious and Violent Offender Re-entry Initiative (SVORI) intervention group had their parole revoked, compared with 21.7% of the control group. This difference was not statistically significant (χ^2 (1, N = 177) = 0.16, p > .05), suggesting that the intervention had little effect on either technical breaches of parole or re-arrest rates during parole. In contrast, Antonio and Crossett (2017; CASP rating strong) found that 79.5% of the intervention group who were re-incarcerated were deemed to have committed a technical breach of parole, compared with 83.9% of the control group. This difference was not statistically significant at the .05 level[1]. However, when focusing on the high-risk group of offenders, it was established that significantly more control group participants were re-incarcerated for a technical breach of parole compared with the intervention group. This difference was statistically significant at the .05 probability level. Similarly, Clark (2015; CASP rating strong) found that a significantly larger percentage of control group participants were returned to prison for a technical breach of parole, when compared with the High-Risk Revocation Reduction (HRRR) programme participants (79% vs 64%, p < .001). Using multi-variate analysis to control for key differences between groups, Clark also found that over time HRRR participation significantly reduced the risk of a new revocation of parole by 28% (HR = 0.715, SE = 0.017, p < .05).

Given the strength of the design of both the Clark and Antonio and Crossett papers, the sample size of the latter paper and some of the shortcomings of the Bouffard and Bergeron paper, these results offer cautious optimism about the ability of targeted intervention programmes to reduce technical breaches, particularly for high-risk offenders.

[1] It is not always possible to report full statistical details as these do not appear in the relevant studies.

Recidivism: Re-arrest

Re-arrest data were used as a recidivism measure by four studies (Bouffard & Bergeron, 2006; Braga et al., 2009; Clark, 2015; Lattimore & Visher, 2009). Using bi-variate analysis, Bouffard and Bergeron (2006; strong) found that 30.9% of the SVORI group were re-arrested having completed parole, compared with 28.3% of the control group. This difference was not statistically significant (χ^2 (1, N = 177) = .100, p > .05), indicating that the SVORI intervention had no impact on post-parole re-arrest rates. While it is acknowledged that this sub-set of the sample could be considered more motivated since they had already completed parole, this was the same for both intervention and control groups. The HRRR intervention group, studied by Clark (strong), demonstrated lower re-arrest rates than the control group (47% vs 53%), although this was not a statistically significant difference (p > .5). Lattimore and Visher (2009; CASP rating moderate) found similarly encouraging results falling short of statistical significance. In these samples, there are concerns that differences in the composition of groups may have influenced the results of bi-variate analysis. Other studies (e.g., Antonio & Crossett, 2016; Friendship et al., 2003) have analysed results separately for different risk groups, making treatment effects for specific cohorts easier to identify. Where groups vary across critical variables, such as risk levels, this technique is useful in identifying nuance within the results.

Using the Cox regression model to estimate the effects of the Boston Re-entry Initiative (BRI) on re-arrest for both any offence and a violent offence, Braga et al. (2009; CASP rating strong) established a significant (p < .01) 32.1% reduction in the subsequent overall arrest hazard rate and a significant (p < .01) 37.1% reduction in the subsequent violent arrest hazard rate. Both results remained statistically significant after key variables were controlled for (any arrest rate, p < .01; violent arrest rate, p < .05). Although re-arrest is considered a good proxy for recidivism (Travis & Visher, 2005), it should be noted that this study only used arrest data from within the state, indicating that any arrests resulting from out of state offending were not

detected in either group. Bouffard and Bergeron (2006; strong) controlled for key differences between the intervention and control groups, including age, gender, most serious offence and Level of Supervision Inventory-Revised score (risk/need assessment tool). They found that participation in the SVORI intervention significantly reduced the risk of re-arrest by approximately 60% when compared to the control group ($B = -0.929$, $SE = 0.462$, $p < .05$), having controlled for group differences. Given that this paper was rated *strong* in the CASP assessment process, these are encouraging results. Using the same procedure as Braga et al. (2009), Clark (2015) found that participating in the HRRR intervention reduced the risk of re-arrest by 26%, when compared with the control group. This difference did not reach statistical significance ($p > .05$).

All four of the studies using re-arrest rates as a proxy for recidivism established either lower re-arrest rates in the intervention group or lower predicted risk of re-arrest in the intervention group. Two of these studies (Bouffard & Bergeron, 2006; Braga et al., 2009) report statistically significant results indicating that the interventions being measured were able to reduce the risk of re-arrest when controlling for key differences between groups. Clark (2015) and Lattimore and Visher (2009), although not reporting statistically significant results, describe lower re-arrest rates in the intervention groups. Clark also reports reductions in the risk of re-arrest rates which may require a larger sample size to reach significance. These are promising results regarding intervention effects on re-arrest rates.

Recidivism: Reconviction

Official reconviction data is the most common proxy for recidivism used in the studies under review. Six of the 10 studies report results in terms of reconvictions (Clark, 2015; Hanson et al., 2004; Hedderman & Sugg, 1996; Nathan et al., 2019; Procter, 1996; Shaul et al., 2016),

although these are not all measured in the same way. Nathan et al. (2019; CASP rating moderate) and Hanson et al. (2004; CASP rating moderate) use officially recorded conviction data but recognise that this also includes reprimands and cautions, which do not always result in a conviction. Most noteworthy is the fact that Procter (1996; CASP rating strong) only reports reconvictions for sexual offences, making it impossible to establish any treatment effect on general recidivism.

Three studies offer results which suggest a positive intervention effect (Clark, 2015; Hedderman & Sugg, 1996; Nathan et al., 2019), although there are concerns around some of these findings. Two studies used a follow-up period of 24 months, and one used a follow-up period of 12 months (Clark, 2015; strong). Hedderman and Sugg (1996; CASP rating moderate) report that the intervention group were reconvicted of a sexual offence at a rate half that of the control group (4.5% vs 9%) and of a violent offence at a rate one-fifth of the control group (4% vs 20%). They also state that the intervention group were reconvicted at a rate lower than predicted using the Offender Group Reconviction Scale (OGRS) algorithm (9% vs 13%), whilst the control group were reconvicted at a rate higher than predicted (29% vs 23%). In addition, none of the 24 offenders who were assessed as having been 'significantly treated' reoffended within the two-year follow-up period. Statistical significance data were not reported, making it difficult to estimate the strength of these results. Although using a different sample of high-risk violent offenders with a likely personality disorder, Nathan et al. (2019; moderate) found similarly encouraging results in their study. Using an intent-to-treat (ITT) approach by examining all cases offered treatment (including those who did not participate), it was established that there were significantly more offences committed by the intervention group than the control group (95% CI = -0.042, 2.334; $p < .05$). However, when only those cases who participated in treatment were analysed, significantly fewer offences were committed by the intervention group compared with the control group (95% CI = -2.014,

-0.096; $p < .05$). It is suspected that the difference in these findings is due to a single case in the ITT group, who committed nine offences (the highest of the entire study population) during the two-year follow up period.

Controlling for a range of relevant variables, Clark (2015; strong) found that HRRR participation reduced the risk of reconviction by almost 42% (HR = 0.581, SE = 0.275, $p < .05$). The relatively low frequency of reconviction and the lack of statistical power in the model (0.64) means that these results need to be interpreted with caution. Collectively these results offer a degree of optimism regarding the ability of intervention programmes to reduce the incidents or risk of reconviction for a new offence, albeit during a relatively short follow-up period. It should also be noted that sample sizes were relatively small, especially in the Nathan et al. (2019; moderate) study, where it is suspected that a single outlier with 9 convictions in the intent-to-treat group skewed the results.

A further three studies offered no support for the hypothesis that community-based interventions can have a positive effect on reconviction rates (Hanson et al., 2004; Procter, 1996; Shaul et al., 2016). Hanson et al. (2004; moderate) found only very small, non-significant differences between the intervention and control groups for sexual reconvictions (21.1% vs 21.8%), violent reconvictions (42.9% vs 44.5%) and all reconvictions (56.6% vs 60.4%), after the 12-year follow-up period. As the full sample was not followed for this length of time, seven-year follow-up data is also provided, indicating the same small, non-significant differences between groups, with sexual and violent recidivism occurring in higher proportions in the intervention group.

Procter (1996; strong) and Shaul et al. (2016; CASP rating moderate) both use survival analysis to establish how long each group remains in the community before reconviction. Neither study found significant differences in survival rates between intervention and control groups. After

12 months, Procter established survival rates (cumulative proportion surviving) of 0.98 for those who had undertaken treatment for sexual offending and 0.94 for the control group. These differences narrowed slightly up to 36 months and then began to widen until the survival rates at 60 months were 0.92 for the intervention group and 0.70 for the control group. Whilst these differences appear positive, they did not reach statistical significance. It should also be noted that as both groups in this study reached the end of their supervision period they were removed from their respective group. At 60 months, this resulted in only 14 intervention and 16 control participants remaining in the community. Similarly, Shaul et al. (2016; moderate) found no effect of the brief Motivation Enhancing Intervention (MEI) on reconviction rates or on time to reoffending.

Results demonstrating the effects of targeted interventions for high-risk violent and sexually violent offenders on reconviction rates are mixed. Some promising results have been established with well-designed studies, often focusing on small samples of complex offenders (e.g., Clark, 2015; Nathan et al., 2019). It is also noteworthy that two of the three studies which provided promising results used a randomised control design, giving the results some additional strength. Two of the three studies focusing on men convicted of a sexual offence provided no support for specific offence focused treatment. The potential implications of these findings are discussed below.

Recidivism: Re-incarceration

Three studies (Antonio & Crossett, 2017; Clark, 2015; Lattimore & Visher, 2009) used re-incarceration as an outcome measure. Both Antonio and Crossett (2017; strong) and Clark (2015; strong) found encouraging results for the effects of both the Cognitive Life Skills (CLS) programme and the High-Risk Revocation Reduction (HRRR) programme (respectively) on

re-incarceration rates, although the latter study fell short of statistical significance. Antonio and Crossett (2017; strong) found significantly lower re-incarceration rates for the CLS group than for the control group across their full sample (23.3% vs 27.3%, $p < .01$) and even greater differences between the groups when looking at the high-risk only sub-sample (23.9% vs 31.7%, $p < .001$). When examining the reasons for re-incarceration, Antonio and Crossett (2017) established that significantly more of the high-risk intervention group were re-incarcerated for a further offence (17.2% vs 11.4%, $p < .05$) and significantly fewer for a technical breach of parole (81.1% vs 87.0%, $p < .05$) when compared with the high-risk control group. Whilst this suggests that the CLS intervention has a positive effect on adherence to parole but either a neutral or negative effect on further offending, additional explanations for these findings are discussed below. Clark established that the intervention group were re-incarcerated at a lower rate than the control group (10% vs 14%, $p > .05$). This result was not statistically significant, perhaps because the sample size was relatively small ($n = 162$), and the frequency of re-incarceration was relatively low.

Both studies controlled for a range of variables and used survival analysis to establish the effects of the intervention on length of time in the community. Antonio and Crossett (2017; strong) found that participation in the CLS reduced the daily probability of re-incarceration by a factor of 0.76, equating to a 24% reduction in returning to prison ($B = -0.28$, $SE = 0.061$, 95% CI = 0.67, 0.85; $p < .001$). The same analysis on the high-risk sub-sample established an even stronger intervention effect equating to a 31% reduction in the risk of re-incarceration ($B = -0.37$, $SE = 0.079$, 95% CI = 0.59, 0.81; $p < .001$). Clark (2015; strong) found that HRRR participation reduced the risk of re-incarceration for a new offence by 34% although this did not reach statistical significance. Caution should be exercised when interpreting these positive but non-significant results because of the low frequency of re-incarceration and the limited statistical power in the model (0.62).

Lattimore and Visher (2009; moderate) found less promising results with no significant differences in re-incarceration rates between the SVORI and control groups up to 24 months (42% vs 39%, $p > .05$). It was concerning that re-incarceration rates were higher for the SVORI group from the six-month follow-up through to the 24-month follow-up, although these differences were also non-significant. This suggests that the SVORI intervention had, at best, no effect on re-incarceration rates and, at worse, a negative effect. However, given that the SVORI group were also found to have been arrested for fewer offences within the 24-month follow-up period (non-significant), these differences could be a result of local enforcement decisions or sentencing policy.

Despite Lattimore and Visher's (2009; moderate) large sample providing no support for the effect of the SVORI intervention on re-incarceration rates, Antonio and Crossett (2017; strong) and Clark (2015; strong) offer promising results for the effects of CLS and HRRR, respectively, on re-incarceration rates. These were well-designed studies, one of which had a large sample size, suggesting that targeted interventions can have a positive effect on reincarceration rates with high-risk, complex offenders.

Access to Community Services

Two studies (Bouffard & Bergeron, 2006; Clark, 2015) report on access to services in the community and Clark specifically analyses the effects of service uptake on recidivism measures. Bouffard and Bergeron (2006; strong) established that the SVORI intervention group were referred to significantly more services in the community than the control group (5.4 services per offender vs 2.4 per offender, $p < .01$). These results were also significant at the same level when examining referrals to specific services, such as chemical dependency aftercare treatment, cognitive programmes, anger management, education services,

parenting/family services and mental health services. When examining participation rates, the uptake of services was proportionately lower in the SVORI group (79.7% vs 89.6%, $p < .05$), although this still led to significantly higher numbers of programmes being participated in by the SVORI group (4.2 services per offender vs 2.2 per offender, $p < .01$), due to the significantly higher number of referrals made.

Using the Cox regression model, Clark (2015; strong) found that for an increase in re-entry services received by both groups there was a corresponding 11% drop in the risk of parole revocation (HR = 0.893, SE = 0.041, $p < .01$) and a 9% drop in the risk of re-arrest (HR = 0.911, SE = 0.048, $p < .05$). There were also positive associations between the receipt of post-release services and reconviction and re-incarceration, although these did not reach significance. Specifically, employment assistance was associated with significant reductions in parole revocation (33%, $p < .05$), re-arrest (39%, $p < .05$) and re-incarceration (68%, $p < .05$) whilst subsidised employment was significantly associated with reductions in parole revocation (57%, $p < .001$), re-arrest (70%, $p < .01$) and reconviction (76%, $p < .05$). When examining the risk of parole revocation, it is notable that, in addition to employment services, a range of other services had a significant positive effect, including enhanced case planning (28% reduction, $p < .05$), community hub participation (35%, $p < .05$), community CBT (32%, $p < .05$) and transportation assistance (35%, $p < .01$). Collectively, these results are encouraging and suggest that specific intervention services can assist those who have previously violated parole to remain in the community for longer.

Housing and Employment Outcome Measures

Lattimore and Visher (2009; moderate) reported housing and employment related outcome measures in addition to recidivism outcomes. They found that the SVORI group were

significantly more likely to report housing independence after 15 months compared with the control group (OR = 1.569, $p < .05$). Differences between the intervention and control groups regarding reported housing stability and having no housing challenges were negligible. Regarding employment, SVORI participants were significantly more likely than the control group to receive formal pay (OR = 1.829, $p < .01$) after 3 months and to have in-job benefits after 3 months (OR = 1.400, $p < .05$), 9 months (OR = 1.602, $p < .01$) and 15 months (OR = 1.386, $p < .05$). SVORI participants were also significantly more likely to report supporting themselves with a job after 15 months (OR = 1.617, $p < .01$). Whilst this indicates some promising effects of SVORI participation on both housing and employment outcomes, these were two of the central aims of the intervention and there are numerous housing and employment measures which show no differences between groups at 3-, 9- and 15-month follow-up stages. Critically, there were no group differences in terms of numbers of months worked and participants that had worked every month since release at any of the follow-up stages. Lattimore and Visher (2009) did not analyse the specific effects of housing and employment outcomes on recidivism measures.

Discussion

Main Findings

Detailed analysis of the ten papers under review revealed mixed but promising results regarding the ability of targeted intervention programmes to have a positive influence on recidivism measures for high-risk, violent and sexually-violent offenders.

With regard to technical parole revocation, there were positive results from two well-designed and high-quality studies (Antonio & Crossett, 2017; Clark, 2015), particularly with high-risk offenders. Given that one study showed that approximately four out of five re-incarcerated

offenders committed technical rather than criminal violations, significant reductions in technical breaches of parole are a welcome finding for this group of complex offenders.

All four studies which reported on re-arrest rates established encouraging results. Clark (2015) and Lattimore and Visher (2009) both found lower re-arrest rates in the groups exposed to the HRRR programme and the SVORI initiative, although these did not reach statistical significance. However, Braga et al. (2009) and Bouffard and Bergeron (2006) both found a significantly lower risk of re-arrest in their Boston Reentry Initiative (BRI) and SVORI cohorts. It is anticipated that those receiving targeted intervention programmes may naturally be monitored more closely due to their engagement with professional agencies supporting high-risk offenders. In this respect, results indicating that intervention groups were re-arrested at lower rates than control groups are particularly encouraging.

The impact of targeted intervention programmes on reconvictions rates for high-risk offenders is less clear. Some studies established an intervention effect on reconviction rates (Clark, 2015; Hedderman & Sugg, 1996; Nathan et al., 2019), although Hedderman and Sugg (1996) reported their results poorly and the remaining two studies contained small or relatively small sample sizes. Hanson et al. (2004), Procter (1996) and Shaul et al. (2016) were unable to establish any positive results. These finding offer little support to the notion that targeted interventions can have a positive impact on reconvictions rates among high-risk, complex offenders.

It was established by Antonio and Crossett (2017) that the CLS programme significantly reduced rates of re-incarceration, particularly among the high-risk sub-sample. Clark (2015) also established lower rates of re-incarceration for the HRRR programme group, although this did not reach statistical significance. Lattimore and Visher (2009) found no effect of SVORI on re-incarceration rates, with the intervention group marginally more likely to return to prison

than the control group. This provides a mixed picture of the effects of interventions on re-incarceration rates, and this is explored in more detail below.

Two separate studies (Bouffard & Bergeron, 2006; Clark, 2015) offer interesting results about the access to and impact of specific services on recidivism. A specific aim of the SVORI project analysed by Bouffard and Bergeron (2006) was to increase access to services and their results suggest that this was achieved. Clark (2015) reports that an increase in access to services had a significant effect in reducing the risk of parole revocation and the risk of re-arrest. Increased access to services also decreased the risk of reconviction and re-incarceration, although this did not reach statistical significance. As only two studies reported on the impact of services on recidivism measures, it is not possible to draw any definitive conclusions in this area.

In terms of housing and employment outcomes, only one study examined these in detail (Lattimore & Visher, 2009). Whilst some results indicated that the intervention group had improved housing and employment outcomes, there were several other measures which showed no differences between groups.

Strengths and Weaknesses of Systematic Literature Review

The search strategy is detailed above and, whilst not exhaustive, it was designed to be as comprehensive as possible and to return a wide range of literature from a variety of disciplines, across different countries. Specific search criteria were arrived at after an extended period of trial and error on relevant databases to familiarise the author with search strategies and terminology. Relevant words or phrases were then modified to ensure that all possible variants would be identified during searches. Database selection and search criteria are considered a strength of this review.

It is recognised that limits had to be placed on searches to ensure the relevance of the results and also for pragmatic reasons related to data management. The decision to exclude books and unpublished dissertations requires acknowledgement. It was felt that the majority of empirical research would be published in peer reviewed journals, although it is accepted that excluding book chapters may have compromised the quantity of research available for review. Whilst including unpublished dissertations may have reduced publication bias, it was expected that they may also have suffered from methodological flaws and, pragmatically, may have been challenging to access. It is also recognised that the reference lists of all relevant articles were scanned using titles, thus falling short of a systematic examination, leading to a potential increase in selection bias. Once searches had been completed, 10% of the final dataset were reviewed by an experienced researcher unconnected to the study to minimise the risk of selection bias.

One of the articles considered for inclusion (Zhang et al., 2003) was unable to be accessed as a full text article. Despite a range of searches being conducted and the primary author being contacted, this was unable to be resolved and the article was excluded from consideration. It is recognised that this may have limited the results of the review.

The Critical Appraisal Skills Programme (CASP; Critical Appraisal Skills Programme, UK, 2018) quality assessment form was selected from a wide range of tools to guide the quality assessment of studies. Whilst it is acknowledged that CASP does not provide a formal scoring process, it allows for a flexible approach whilst alerting the user to the design and methodological considerations required for a final assessment. This also allowed for a bespoke rating system to be developed to assist with the final assessment decision. It is recognised that the flexible structure of this tool can be considered both a weakness and a strength of the current review.

Inclusion criteria stipulated that studies needed to focus on exclusively, or majority, male cohorts. This decision was made to ensure the inclusion of all relevant results. Several of the studies included a minority of female participants and, whilst some of these studies reported results by gender (e.g., Lattimore & Visher, 2009), several did not (e.g., Antonio & Crossett, 2017), making it difficult to assess the impact of including female offenders in the total sample. As the aim of this review is to examine the impact of community-based interventions on high-risk, male offenders released from custody, this is a drawback of the systematic review process.

30% of the included studies focused exclusively on men convicted of a sexual offence, whilst the remaining 70% focused on high-risk, violent, sexually-violent or repeat parole violators. Whilst these offenders were generally considered high-risk, using a variety of measures, they represented different groups of offenders who may be expected to respond differently to interventions. In addition, the measures used to assess risk levels differed across countries, making it hard to guarantee the comparability of cohorts based on risk. Finally, the personality profile of cohorts was rarely specifically measured, instead offence typology and risk level were used as proxies for personality complexity. In combination, these factors will inevitably create variability across cohorts, making comparisons and generalisations challenging.

Strengths and Weaknesses of Studies Under Review

60% of the studies reviewed used North American cohorts, a further 30% used UK cohorts and the final study was based in the Netherlands. This bias towards western, developed, economies makes it challenging to generalise results to other offending populations. This is particularly the case where criminal justice systems have different thresholds for recall, arrest, conviction, and incarceration, as all the studies included in this review used at least one of these measures as a proxy for recidivism.

Naturally there were variations in follow-up periods, although some studies appeared not to follow participants for long enough to detect potential outcomes. Hedderman and Sugg (1996) followed up a group of men convicted of a sexual offence for two years post-intervention. Given the base rate of sexual offending is relatively low (Jung et al., 2018; Nicholaichuk et al., 2014) the authors recognise that this was a drawback of their study. Other follow-up periods were longer for studies focusing on men convicted of a sexual offence, although Procter's 5-year follow-up suffered from methodological issues, as those who had completed their licence during this period were removed from the study. This meant that an offence committed after licence completion but within the 5-year follow-up period was not counted as recidivism. Although follow-up periods varied both between and within studies, differences in *time at risk* tended to be controlled for in statistical analysis, making this a strength of most of the studies reviewed.

Due to the way that analysis was conducted, it was rarely possible to determine how sub-groups of participants responded to interventions. Antonio and Crossett (2017) were able to determine that as age increased there was a significant reduction in the risk of re-incarceration, although this analysis was not possible in most other studies. This prevents the review from drawing more nuanced conclusions about "what works for certain offenders under what conditions?" (Antonio & Crossett, 2017, p.515).

Recidivism measures varied significantly across the ten studies, with revocation of parole, re-arrest, reconviction and re-incarceration all being used as proxies. Whilst this made comparisons between studies more challenging, it also allowed for a better understanding of the strengths and weaknesses of each measure. Six of the studies reviewed used a single measure of recidivism, with the remaining four using multiple measures. Only Clark (2015) used all four available proxies for recidivism. The use of multiple measures of recidivism in some studies is considered a strength, as it allows the reader to understand the impact that an

intervention may have on different forms of recidivism. For the same reasons, the use of single recidivism measures in several studies is considered a weakness.

A particular strength of three of the studies reviewed (Clark, 2015; Nathan et al., 2019; Shaul et al., 2016) is that they used a randomised control trial (RCT) design. This significantly reduces the risk of bias in the assignment of participants to groups and therefore adds confidence when interpreting results. In addition, all three studies used intent-to-treat (ITT) analysis, thus offering an unbiased exploration of the efficacy of their respective interventions based on initial randomisation. Although a potential weakness of using ITT analysis is the dilution effect of those not receiving treatment being included in the intervention cohort (Gupta, 2011), two of the studies (Clark, 2015; Nathan et al. 2019) also used per-protocol or as-treated analysis to minimise this effect.

Interpretation of Findings

This review confirms that it is possible to run interventions and design research which demonstrates a positive effect on different measures of recidivism. Focusing on those studies which established positive effects of interventions, the majority included a custodial element. Whilst this review focused on community-based interventions, it did not exclude those studies which incorporated a transitional process, beginning in custody. Braga et al. (2009), Bouffard and Bergeron (2006), Clark (2015), Lattimore and Visher (2009) and Nathan et al. (2019) all found either statistically significant or promising effects of interventions on recidivism and all included elements of custodial contact with participants prior to release. It is acknowledged that for Nathan et al. (2019) this involved interviews and assessments and for Clark (2015) this part of the intervention was not consistently delivered. It is also recognised that Antonio and Crossett (2017) and Hedderman and Sugg (1996) also established positive results but with no

custodial element to the study. However, these finding suggest that supporting high-risk complex offenders in their transition from custody to the community may have a positive effect on recidivism directly or on their engagement with services aimed at reducing recidivism. This finding is consistent with Berghuis (2018) who established that continuity of care from custody to community was a key feature of successful intervention programmes.

It is clear from those studies which used multiple measures of recidivism, that interventions can have different effects on each measure. This review draws out these differences and highlights one point in particular: Revocation of parole was used as a measure in three studies (Antonio & Crossett, 2017; Bouffard & Bergeron, 2006; Clark, 2015) and appears to differ significantly from other officially recorded measures of recidivism. According to Antonio and Crossett (2017), approximately 80% of parole breaches were for technical, as opposed to criminal, violations. This suggests that revocation of parole is a better proxy for adherence to parole than it is for recidivism. The offenders studied in this review are either explicitly described as likely to have a personality disorder (e.g., Nathan et al., 2019) or are selected for inclusion as they display functioning closely aligned with features of personality disorder, particularly anti-social personality disorder (ASPD; e.g., Braga et al., 2009). One of the features of ASPD is a '*failure to conform to social norms with respect to lawful behaviours*' (HMPPS & NHS, 2020), making it extremely challenging for such offenders to adhere to their parole conditions. Those studies which indicate a positive effect of an intervention on parole revocation are perhaps measuring the ability to work collaboratively with professionals, as opposed to recidivism. It should also be considered that Probation services tend to have the most control over parole violation decisions, indicating that those being supported through high-intensity interventions may be afforded more latitude than other released offenders. In conclusion, it appears important to separate studies measuring revocation of parole from those measuring recidivism, as they are different outcomes affected by different processes.

Whilst those supported through high-intensity interventions may be championed by Probation staff, they may also be more likely to come to the attention of law enforcement. The interventions studied by the current review most often involved engagement with numerous community services. Whilst not the intention of these interventions, this inevitably provides more opportunity for monitoring, increasing the likelihood of revocation, re-arrest, reconviction and re-incarceration. This highlights the fact that measures such as reconviction are an undercount of offending (Hedderman & Sugg, 1996) and perhaps suggests that those more intensively supported offenders are more likely to be apprehended than those managed by standard weekly or monthly Probation supervision.

Although Hedderman and Sugg (1996) established some positive results of community-based treatment for men convicted of a sexual offence, the three studies focusing on this cohort generally found limited or no treatment effect. These studies are also relatively old, with two approaching 25 years since publication. More recent studies (e.g., Antonio & Crossett, 2017; Clark, 2015; Nathan et al., 2019) have taken the different approach of treating violent and sexually-violent offenders through the same pathway. Nathan et al. (2019) adopted a relational model of engagement and support, focusing on the individual's functioning as opposed to their offending history. These more recent studies have also provided more positive results. Whilst it is acknowledged that research design and statistical modelling have progressed significantly during recent years, these findings suggest that interventions focused on the person, as opposed to the offence, may result in more successful outcomes in the future.

Conclusions and Recommendations for Practice

There is a paradox in undertaking research in this field. Since the advent of the Offender Personality Disorder Pathway (OPDP) in the UK, there has been additional focus on the

relational aspect of reducing recidivism, particularly with violent offenders with complex personality difficulties (Ryan et al., 2019). The delivery of relational interventions, such as that outlined in the Nathan et al. (2019) paper, mean that they are most effectively delivered on a small scale over an extended period. Studies focusing on these interventions are then criticised for having small sample sizes and limited statistical power. Conversely, studies involving large cohorts, such as those outlined in the Antonio and Crossett (2017) and Lattimore and Visher (2009) papers, are criticised for the difficulties they experience in individualising services and the lack of nuance reported in the results. It is suggested that there needs to be a balance struck between the delivery of small-scale, relational-focused interventions and larger-scale skills-based interventions. This would enable those whose offending is functionally related to their personality to engage in high-intensity services to support their resettlement. Likewise, those who may benefit more immediately from education, employment and housing support can be assisted via a different route. This distinction relies on detailed analysis of the needs of individuals released from prison to direct them to appropriate services.

Findings suggest that services supporting men released from prison should be mindful of incorporating transitional support. In practice, this refers to the building of professional relationships with service users prior to their release to minimise the distress experienced during transition and to increase the chances of an extended period in the community. Although this review was unable to offer specific findings in this area, attention should be paid to the type of transitional services on offer, as recent literature suggests that a relational model of support, both in and out of prison, can work well for the most complex offenders (Joseph & Benefield, 2012; Ramsden et al., 2017; Ryan et al., 2019). This is discussed further in Chapter 5.

Regarding future research in this area, it may be prudent to consider what type of service user we are interested in studying. Large-scale research is a useful way of establishing general intervention effects, although it is often challenging to fully understand the nuance within the data. Alongside research using large cohorts, it is recommended that future research aims to replicate small-scale studies of services providing individualised interventions to high-risk released offenders. In addition, research should aim to draw together data from a range of small-scale studies over a long enough period to allow for analysis of larger participant numbers. This type of research would enable practitioners to understand more about who specific services work best for. The current review was able to provide some analysis of intervention effects according to age, but further data is required before conclusions can start to be drawn. In addition, it is important to understand potential differences in intervention effects according to personality functioning, history of parole revocation and quality of family and social support, among other variables. This will significantly assist the process of assigning individuals to appropriate interventions.

The cautiously optimistic results offered by the current review contrast with those outlined by Berghuis (2018). Although including lower-risk offenders and exclusively US cohorts, the sole use of RCT studies in the Berghuis review is noteworthy. It is possible that more rigorous design prevents false-positive intervention effects from being detected, although it is equally possible that attrition rates and other complexities with this group prevent RCTs from being effectively implemented. It is recommended that further research uses both RCT and cohort designs to add to our understanding in this area.

Finally, this review highlights the need to separate studies using revocation of parole and recidivism as outcome measures. As is documented above, approximately 80% of parole revocations are for technical breaches, rather than criminal behaviour. This suggests that more attention needs to be paid to interventions aimed at keeping released offenders in the

community for longer. For those men and women whose identities are tied to their criminal associates and their experiences in prison, extended time in the community offers them the opportunity to develop a different identity, benefitting themselves and society more generally. Furthermore, consideration should be given to moving away from recidivism-based outcome measures and towards reintegration measures, which are less likely to be influenced by legal and political decision making.

This literature review provides the context for and has informed the design of the empirical study in the next chapter, which focuses on the same population. Whilst the study design was initially influenced by the availability of high-quality data, a variety of amendments were made on the basis of the systematic literature review. The importance of comparing groups who were as similar as possible across a range of key variable was established and the current literature review added to the understanding of what these variables are. Strengths of a number of the studies examined above were also built into the design of the empirical study, including the use of propensity score methods to create well matched intervention and control groups and the use of multiple measures of recidivism. Where possible, weaknesses of the examined studies were also avoided in the following empirical study, for example by ensuring that analysis of outcomes for sub-groups of each population (e.g., by age and ethnicity) was conducted. The empirical study which follows in the next chapter develops our understanding of the efficacy of existing intervention services for high-risk, violent, complex men and examines the variables which affect survival times in the community. The study also contributes to the ongoing debate about how cases are selected for interventions and how success is accurately measured.

CHAPTER 3

Empirical Research

Are Community Intervention Services Effective at Keeping High-Risk, Complex, Violent, Male Offenders in the Community for Longer?

Introduction

Management of Complex Violent Offenders

In 2021, 47,014 prisoners were released from UK prisons into the community, with 21,214 being recalled to custody following a previous release (HMPPS, 2022). A further proportion of these released offenders were returned to custody following reconviction, although separate data on recalls and reconvictions are not currently available. 68% of those recalled to custody in 2021 had either served a determinate sentence of 12 months or more or an indeterminate sentence (HMPPS, 2022), suggesting that over two-thirds of those recalled had been convicted of a serious offence. The current study focuses on men convicted of violent or sexually violent offences who are considered high-risk and complex, due to personality difficulties. Not only are return to custody rates high for this group (e.g., Polaschek *et al*, 2018), but there are also significant social and economic issues that result from the reincarceration of this group, discussed earlier in Chapter 1. The majority of those committing violent or sexually violent offences and considered high-risk also experience traits of personality disorder (Davison & Janca, 2012; Stone, 2007). It is known that there are high rates of reoffending and recall among male offenders with personality disorder (Bruce et al, 2017; Coid et al, 2006), with Fridell et al (2008) finding that those with a diagnosis of personality disorder were twice as likely to recidivate than those without a diagnosis. Offenders with personality disorder are also known to have experienced a disproportionate number of adverse childhood experiences (Fox et al., 2015), which have been shown to be linked to violent recidivism (Craig & Zettler, 2021). This suggests that reintegration into the community for high-risk, violent, personality disordered offenders is more complex and prone to failure than for other offending groups.

Methodological Issues with Previous Research

A variety of approaches have been taken to manage this group of released offenders in the community, generally split between external monitoring/controls (e.g., Duwe & McNeeley, 2021), risk related treatment (e.g., for substance misuse or personality traits) or lifestyle support, such as housing, employment, and social support (e.g., Rydberg et al, 2022). Research focused on the efficacy of such post-release community intervention services is challenging for a range of reasons. There are numerous interpretations of recidivism, including re-arrest for a new offence, re-conviction for a new offence or a return to custody for a technical breach (Braga et al., 2009). A commonly used measure in recidivism studies is official reconviction rates (e.g., Nathan et al., 2019; Shaul et al., 2016), however, these are subject to variations in local policing and court and probation policies across both place and time (Travis & Visher, 2005). In addition, it is estimated that up to four in every five offenders reincarcerated whilst on licence are returned for a technical breach (Antonio & Crossett, 2017), such as disengaging from supervision, as opposed to being convicted for a further offence, thus creating a distorted impression. Recidivism research often uses different follow-up periods (Fazel & Wolf, 2015), making conclusions about successful releases hard to reach. Within the follow-up period, most studies do not distinguish offenders who reoffend after two days from those who reoffend after two years (Peters et al., 2015). Whilst there are methodological reasons why this form of research is easier to conduct, it is noteworthy that previous research has attempted to identify interventions which *prevent* reoffending rather than *delay* reoffending. It is argued that interventions that can prolong release in the community should be considered successful for several reasons. The longer released offenders spend in the community the less institutionalised they are likely to be, the more likely they are to find appropriate housing and employment and the longer the potential delay to reincarceration, thus reducing the cost to society.

Although various studies have established positive effects of community-based treatment using a range of outcome measures (e.g., Antonio & Crossett, 2017; Bouffard & Bergeron, 2006; Braga et al., 2009; Clark, 2015), many of these experience issues with selection bias (Peters et al., 2015), as it is often the case that those factors which influence propensity for treatment success will also influence the likelihood of recidivism. This means that cases selected for and agreeing to engage in treatment may be more likely to remain offence free, regardless of treatment efficacy, than those not selected or refusing to engage in treatment. To minimise the risk of selection bias, propensity score methods have been increasingly used in forensic research over recent decades (Jolliffe & Hedderman, 2012). Propensity score methods are a collection of techniques which aim to make intervention and control groups as similar as possible across a range of identified covariates. Such methods are used in observational research to control for the effects of case selection on the outcome measure. However, caution is required when interpreting propensity score methods in research, as Thoemmes and Kim (2011) report in their systematic review that many studies use an inappropriate method to minimise selection bias and do not report sufficient detail to allow an understanding of analytical choices, such as which specific method was used, which covariates were matched or whether all intervention cases were included in the matching process.

Current Management of High-Risk Offenders in England and Wales

Over recent years, there has been a significant increase in services for high-risk, complex, violent offenders released into the community, although implementation has been sporadic. This means that offenders classified as high-risk and complex, regarding personality traits, have been managed differently according to the services available in their area. Whilst this is far from ideal, it does mean that research into the efficacy of different types of interventions

for this group of offenders is possible, as groups of comparable community-based offenders have been managed by low, medium, and high intensity interventions over the same time period. Descriptions of some of these services are provided below.

Intensive Intervention and Risk Management Services (IIRMS)

Since funding for the Dangerous and Severe Personality Disorder (DSPD) programme (Home Office & Department of Health, 1999) was diverted to the creation of the Offender Personality Disorder Pathway (OPDP; Joseph & Benefield, 2012), an equal focus has been placed on custodial and community management of complex men and women experiencing personality difficulties (Bruce et al., 2017). Although two intensive community risk management services have been in existence for over 20 years and three further services for over ten years, they have more recently been incorporated into the OPDP strategy and numerous further services have been commissioned (OPDP, 2018). These have been branded Intensive Intervention and Risk Management Services (IIRMS) and are designed to be flexible in their implementation, whilst aligning to four underpinning theoretical domains: attachment theory (e.g., Bowlby, 1969); desistance (e.g., Maruna, 2001) and the Good Lives Model (GLM; Ward & Stewart, 2003); theory of needs (e.g., Deci & Ryan, 2000); and the Risk, Need, Responsivity model (RNR; Andrews et al., 1990). It is acknowledged that the GLM and RNR models contain overlapping principles, albeit approaching the issue of criminal recidivism from opposite perspectives (Andrews et al., 2011). Whilst services are not prescriptive, to prevent the stifling of innovation, they all adhere to these basic principles when engaging and supporting service users in the community.

Service users are deemed suitable for IIRMS if they have been identified as having traits of *severe* personality disorder, more suitably termed *complex emotional needs*. They also need to

be assessed as presenting a high risk of harm, with a functional link between personality traits and offending (Baker et al., 2013). Whilst these men and women are considered some of the most complex cases managed by the Probation Service, they are deemed manageable in the community or have been released from determinate sentences. The IIRMS specification (HMPPS & NHS, 2021) stipulates an in-reach phase, an induction and intervention phase and

Figure 2: Typical IIRMS Process

WHO IS REFERRED: A typical IIRMS case is a high risk, violent or sexually violent offender who screens into the Offender Personality Disorder Pathway. They will often have recently been released from custody or have become unstable after a period of time in the community.

WHAT IS DELIVERED: The interventions offered are bespoke, making it difficult to provide an example intervention. However, they will be formulation led, meaning that a full understanding of what led a person to their current situation is developed and then decisions are made about which treatment is likely to stabilise or reduce risk most effectively. Among others, this can include psychoeducation work, practical support, risk reduction strategies, and therapeutic interventions.

THERAPEUTIC MODALITY: Where therapy is offered, risk reduction work can include CBT, DBT, schema, or trauma focused approaches, among other modalities.

WHO DELIVERS: IIRMS interventions are typically offered by qualified psychologists or trainee psychologists under supervision. They can also be delivered by specialist Probation Officers who are part of the OPDP team. Delivery staff receive supervision and support from qualified colleagues within OPDP.

LONGEVITY AND FREQUENCY: This is a formulation led decision, but interventions are typically between 3 and 12 months, depending on need, and frequency is usually weekly.

MEARUREMENT: Each IIRMS initially had its own method of monitoring progress which typically included a wellbeing and a risk measure. Current IIRMS now use the CORE-10 (Barkham et al., 2012), Adult Hope Scale (Snyder et al., 1991), and the Social Functioning Questionnaire (Tyrer et al., 1985).

an exit phase for all participants (Ramsden et al., 2017), meaning that service users are supported through their transition from custody to the community. These services are non-residential, and the specific intervention is then guided by a formulation, which provides a psychological explanation of the underlying mechanism of the presenting problem and offers suggestions for change (Logan & Johnson, 2010). Figure 2 outlines a typical IIRMS process, although it should be recognised that it is not possible to provide a case study, as IIRMS is designed to be bespoke, catering for the individual needs of each service user.

Intensive Response Intensive Service (IRiS) Intervention

Some Probation areas have developed intervention services through the Integrated Offender Management (IOM) model (Ministry of Justice, 2015). Such services offer additional support and monitoring to cases deemed high-risk and who may experience complex personality traits or have prolific offending histories. One of these services, Integrated Response Integrated Service (IRiS), is included in the current study and is described below.

IRiS has been operating in a mid-sized English city and surrounding rural county since 2012 (IRiS, 2020). As part of the IOM model, it is a non-residential, co-located service involving

Police, Probation Service (PS) and Offender Personality Disorder Pathway (OPDP) staff. The aim of IRiS is to reduce reoffending by developing collaborative working practices using multi-agency input, relying on evidence-based rehabilitation strategies, and balancing psychological approaches with Police enforcement. The cohort managed through IRiS are adults considered to pose a high risk of harm to the public, children, or known victims. IRiS is a formulation led service that offers enhanced monitoring, psychological guidance to Probation Practitioners and, dependent on need, referrals to bespoke treatment pathways delivered by specialist PS staff, OPDP psychologists or other Health / OPDP teams. Most cases managed by the IRiS service also screen into the OPDP, although this is not a prerequisite for acceptance. However, to ensure that cases were as comparable as possible, the current study only used IRiS cases that screened into the OPDP. Figure 3 outlines a typical IRiS process, although it is not possible to be more specific due to the bespoke nature of the intervention service.

Figure 3: Typical IRiS Process

WHO IS REFERRED: Adults who are considered to pose a high risk of harm to the general public or known others. Although there is no requirement to screen into OPDP, cases are typically complex, often experiencing difficulties with mental health, personality disorder or substances.

RISK MANAGEMENT PLANS (RMPs): Multiagency RMPs are developed for each accepted referral based on the four pillars of offender management: Supervision; Monitoring and Control; Intervention and Treatment; and Victim Safety.

HOW ARE CASES MANAGED: Accepted cases are assigned both a Probation Offender Manager and a Police Offender Manager, who work jointly to manage the case. Cases are supervised regularly, and supervision is guided by a bespoke formulation. Intelligence is also

monitored across agencies. If risk is deemed to be increasing, scrutiny is increased, and strategies are used to disrupt potential offending behaviour.

JOINT WORKING: Cases can be jointly worked with members of the OPDP team, who offer a maximum of 6 sessions to co-work the case to resolve a specific difficulty.

LONGEVITY AND FREQUENCY: Cases will typically be required to attend Probation supervision weekly and may also be required to meet their Police Offender Manager weekly.

MEARUREMENT: Recidivism measures are used to assess progress.

Probation as Usual

As specific intervention services for high-risk, complex, offenders released into the community have not been routinely available across all areas, the default process is to be managed under licence by the Probation Service (PS). Management under licence involves regular meetings with Probation Practitioners and adherence to a series of bespoke licence conditions, dependent on the specific risk and need of the individual (Ministry of Justice and HM Prison and Probation Service, 2022). In addition, many high-risk, complex, offenders are required to reside in an Approved Premise (AP) managed by PS staff. Each AP has a set of rules which residents are required to abide by, including bespoke sign-in times and night-time curfews. Some of these individuals are required to engage in substance misuse or offending behaviour programmes in the community, although most high-risk, complex, offenders are required to complete any necessary treatment prior to release in custody.

Current Study

The current study used recall to custody and reconviction data as outcome measures to assess the efficacy of IIRMS and IRiS interventions by comparing them with appropriately matched control groups for high-risk, complex, violent offenders. All cases used in the current study are male and screen into the OPDP, making them high-risk, complex, violent offenders. Using matching techniques accounting for a range of demographic, conviction, risk, and personality variables, separate control groups were identified for the IIRMS, IRiS, and combined intervention groups. The combined group, referred to hereafter as the Intervention group, is an amalgamation of the IIRMS and IRiS cohorts and was examined separately to establish whether any intervention is effective in prolonging released offenders' time in the community, when compared to a 'Probation as usual' control group (see above). The study used survival analysis to determine the effect of each intervention on the length of time released offenders remained in the community. Furthermore, the study aims to examine whether demographic, conviction, risk, and personality variables affect survival rates in the community when combining each intervention group with their matched control group.

Research Questions

The current study aims to answer the following research questions:

- Do IIRMS or IRiS participation influence survival time in the community for high-risk, complex, violent, male offenders when compared to matched Probation as Usual cohorts?
- Which demographic, conviction, risk, and personality variables are predictive of survival time in the community for this population?

Methodology

Ethics

The current study used secondary data jointly owned by the Offender Personality Disorder Pathway (OPDP) and the Ministry of Justice (MoJ) and managed by the North of England Commissioning Support Unit (NECSU). Ethical approval for the current study was granted by the University of Birmingham's Research Ethics Team and by the National Research Committee (Ref: 2019-119). Due to the sensitivity of this research, data were stored on a Ministry of Justice server throughout and access to data was only granted to the primary researcher. Access to identifiable cases was permitted for the primary researcher to examine and cleanse the initial data set. Once this stage was complete, cases were anonymised, after which it was not possible to identify individual cases. Only aggregated data were shared with persons outside MoJ that were connected to the research.

Data Access

All cases used in the current study screen into the Offender Personality Disorder Pathway (OPDP). As the OPDP is a joint MoJ and National Health Service (NHS) funded project, the data for the current study are held by both organisations. The MoJ provided data in relation to demographics, offending, risk levels, and outcomes. The NHS provided data in relation to personality disorder traits and interventions. To access both MoJ and NHS data, a Memorandum of Understanding (MoU) was agreed between the lead researcher and the OPDP Research & Data Team, who have access to both organisations' data. The MoU was signed by both parties on 10th May 2021 (see Appendix G) and was signed off via email by the Information Assurance Team within MoJ on 17th June 2021.

The current study focuses on OPDP cases who were released from custody into the community between 1st April 2014 and 31st March 2019. Starting the current study prior to April 2014 risked excluding several services, as they were either not in existence or were still setting up. Starting the current study after April 2014 meant that valuable data would have been omitted from the study. March 2019 was chosen as the end date for the study as this maximised the available data, whilst allowing for a reasonable follow-up period prior to the start of the Coronavirus pandemic, which affected the community management of the population under review. This is explained further below. For information on high intensity services, data were requested from five Intensive Intervention and Risk Management Services (IIRMS) across England. These services were selected as they had been running for the entire period under review, are considered comparable in both intervention model and intensity and are non-residential in nature. Data were also requested from the OPDP team involved in the delivery of the IRiS project in Bristol. All services selected for the current study are located in a large or mid-sized English city with surrounding rural areas, making them geographically comparable. Only OPDP screened-in cases who were released from custody between the study dates were requested. To compare the IIRMS and IRiS groups with a matched control sample, data were requested from MoJ and NHS for all OPDP screened-in cases released from custody within the study dates across the Midlands Local Delivery Unit (LDU). The Midlands was selected as a control population as there were no IIRMS or IOM led intervention services being run across the region throughout the study period. This request returned over four thousand cases which were then used in the propensity score matching process, described below.

Data Variables

The data variables requested from NECSU for all three populations (IIRMS, IRiS and Control) can be viewed in Table H1 in Appendix H. A list of cases who had engaged in either IIRMS or IRiS interventions during the study dates were requested from individual services. Identifiers for these cases were then used to request the data in Table H1 from NECSU. In addition to these cases, information was also requested on all cases who screened into the Offender Personality Disorder Pathway (OPDP) service who had been managed during the study dates within the Midlands LDU. Only data on cases with at least 1 year left on licence were request.

Data were also requested from IIRMS and IRiS on pre- and post-intervention outcome measures used during the study period, including the CORE-OM where this was used. Unfortunately, the quality and consistency of data was not good enough to be used in the current study. This is largely because services were collecting their own outcome measures, many of which differed between services. Some services did not collect outcome data from the outset and many cases only had pre-intervention measures available, as they disengaged, were returned to custody or did not complete post-intervention outcome measures for other reasons. This meant it was not possible to use outcome measures as a proxy for engagement or as an additional variable with which to measure the efficacy of services.

The data requested fell broadly within two categories: matching variables and outcome variables. The data used to create equitable treatment and control groups related to variables known to have an influence on potential recidivism. These included demographic factors, offence variables, risk levels and additional complexity measures (OPDP screening variables). The outcome variables requested included Probation warning data, licence breach data, official recall data and official reconviction data. Although the intention was to use Probation warning

and licence breach data as additional outcome measures, this was not possible due to the quality of the data. Once the data were received, a quality assessment process revealed that warning and breach information was only sporadically recorded on the data collection system. Unlike recalls and reconvictions, warnings and breaches are not mandatory fields for data collection, meaning that a significant proportion of cases that were warned or breached had no record of such events. Due to concerns around poor quality data, a decision was made not to include warning and breach data in the final study.

Where relevant, data were also requested on the type of offence cases were reconvicted of. The purpose of requesting this data was to examine if there was a harm reduction effect from either of the interventions being investigated. As is mentioned above, the current thesis is interested in moving beyond binary assessments of whether cases *succeed* or *fail* in the community by assessing both how long they survive in the community and also whether there is a harm reduction element to services. Harm reduction refers to the notion of violent offenders reducing the severity of any future offending. Unfortunately, the data provided were not sufficient to complete an in-depth and accurate assessment of any potential harm reduction effects. Whilst the data system holds information on all convictions, it was not possible to link a particular return to custody event with a specific reconviction. Whilst efforts were made to perform this process manually, it was neither accurate enough nor practical to complete this process on over 4,000 cases. This is discussed further below, with regard to recommendations for future research.

Data Cleansing

Prior to matching, the data were examined, and decisions made regarding missing data, duplicate data, and other anomalies. At this stage, only the covariates to be used for matching

were examined. The total number of cases initially received was 4,567, which included 434 IIRMS cases, 107 IRiS cases and 4,026 Control cases. 61 cases were removed from the IIRMS group and 32 were removed from the IRiS group, as they did not fit the study criteria. These cases either did not receive a custodial sentence, were released either before or after the study period, were released but never worked with the relevant intervention service, were deported immediately on release or were duplicate cases. 268 cases were removed from the Control group, either because it was not possible to match the cases to a specific custodial offence or because they were duplicate cases. Before groups were matched, this left an available data set of 4,206 cases, comprising 373 IIRMS cases, 75 IRiS cases and 3,758 Control cases.

For each case, a start date was required to determine the commencement of the *at risk* period. For most cases, including all Control group cases, the start date was the first date a case was released from custody after the index offence. For IIRMS and IRiS cases, the start date was the first date they were accepted into the relevant service. If cases were still in custody at this time, their next release date whilst still working with the relevant service was used as the start date. This decision was made because cases working with a service whilst still in custody are not at risk of recall and are at significantly lower risk of reconviction, although it is acknowledged that reoffending can still occur in custody. It should also be noted that whilst many IIRMS and IRiS cases began work with the relevant service either in custody or shortly after release, some cases had been in the community for several months by the time they started with the relevant service. This is commented on further in the discussion below.

Propensity Score Methods

Treatment allocation for forensic clients is often influenced by baseline characteristics, such as those relating to risk, offence history, personality traits and even demographic information.

Propensity score methods (PSM; Rosenbaum & Rubin, 1983) aim to account for these potentially confounding variables by using them to carefully match intervention and control groups, thus allowing research to more confidently infer causality (Zakrison et al., 2018). Propensity score methods are increasingly being used in both medical and social science research to estimate the effects of treatment in observational studies (Austin, 2014; Diamond & Sekhon, 2013). Whilst it is traditionally accepted that randomised control trials (RCTs) are the preferred research design when estimating the effects of treatment (Austin, 2011; Rubin, 2007), one meta-analytic study found no difference in the estimated treatment effect size between RCT and PSM studies (Olmos & Govindasamy, 2014). It should also be noted that whilst randomisation will balance covariates across treated and untreated samples in general, there is no guarantee it will do this in one specific randomisation (Austin et al., 2010). There are also practical and ethical concerns when using an RCT design in forensic research, particularly with high-risk, violent, and sexually violent offenders in the community (Bonell et al., 2011; Marshall & Marshall, 2007; Ward & Willis, 2010). Observational studies are common in this area of research but suffer from selection bias associated with deciding which individuals are offered an intervention and which are not (Caliendo & Kopeinig, 2008).

A propensity score is the likelihood that a case will be assigned to or selected for a treatment, based on a predetermined set of covariates (Thoemmes & Kim, 2011). This score can then be used to identify control cases which appear most similar across these covariates, thus minimising confounding caused by inherent differences in baseline characteristics between treatment and control groups. It is also possible to use other estimates of difference between treatment and control groups, as well as applying these techniques in a variety of ways to approximate the covariate balance expected from an RCT approach. A full description of the considerations made prior to choosing the propensity score method for the current study, as well as decisions about which matching covariates were selected, can be found in Appendix H.

These decisions are described in detail, as it is essential that the methodology is fully explained to allow for validation and replication (Ali et al., 2015; Shadish, 2013). Table H1 shows the covariates used in a sample of relevant literature, as well as the covariates used in the current study. The same appendix outlines which variables contribute to the creation of the complexity score.

Choosing a Matching Method

In two systematic reviews of the use of propensity score methods across social sciences and medical literature, Gayat et al. (2010) and Thoemmes and Kim (2011) established that reporting of key methodological processes was poor, often preventing the replication of studies. Ho et al. (2007) recognise that the lack of transparency in methodological reporting can reduce reader confidence in the selected method of covariate balance. The current study explores a range of matching methods and fully reports the final method used. A range of balance statistics, both tabulated and graphical, are also reported to assess the final accuracy of covariate balance (Appendices H and I). The statistical package '*R*' was used for the propensity score matching process.

As both the IIRMS and IRiS groups were drawn from different populations, a decision was made to create multiple control groups. This ensures that each intervention group is compared to a control group most closely matching the individuals offered each specific intervention. Although the IIRMS and IRiS groups were combined to create the Intervention group, a decision was made not to simply combine the associated control groups for this population. Instead, the same matching process was conducted for the combined Intervention group to ensure an accurately matched control group for this population. The process of determining the most effective matching technique was run separately for each group, as there is no

guarantee that a method that results in a balanced control group for one intervention group will provide a similarly balanced control group for another intervention group. This has resulted in three separate processes to identify the most appropriate matching method for each population.

The covariate balance between groups was assessed using combinations of nearest neighbour, optimal and genetic matching methods, varying the 'distance' measure between the propensity score and the Mahalanobis distance, using mixed methods to match exactly on specific risk and complexity variables and matching using 1:1, 1:2, 1:3 and 1:4 ratios of intervention to control groups. The Mahalanobis distance, as an alternative to using the propensity score, is an algorithm which calculates the distance between multiple variables (Olmos & Govindasamy, 2015) and can be an accurate way of creating balance when using relatively few variables (Stuart, 2010). Systematically testing the balance between treatment and control groups using the variations described above resulted in 48 potential methods per group. Balance statistics and visual representations were then used to decide on a specific matching method for each group.

Achieving good balance between treatment and control groups requires reductions in bias and variance within the samples. However, by aggressively reducing bias and variance, thus creating a balanced control group, there is an inevitable reduction in the Effective Sample Size (ESS; Golinelli et al., 2012). This results in the potential issue of not having enough statistical power to detect significant correlations in the data. In the current study, using 1:1 matching in the IIRMS group would reduce the potential sample size by 82% (4,131 → 746), weakening the ability of the data to detect significant correlations. As the potential control sample is approximately ten times the size of any of the intervention groups, a decision was made to trial the matching of numerous control cases to each intervention case, referred to as one-to-many (*1:many)* matching. Whilst this increased the available sample size, care was taken to ensure that the balance statistics described in Appendix J remained within accepted tolerances. Care

was also taken to ensure that all intervention cases were both included in the matching process and were also matched to the specified number of control cases, to avoid introducing bias due to incomplete matching (Austin & Cafri, 2020).

Final Matching Methods

Considering the need for both balanced groups and a good ESS, a 1:4 ratio of matching (intervention:control) was used for all three intervention groups. Different matching methods were used for the IIRMS group compared with the IRiS and Intervention groups as they provided the best balance statistics for each group.

Appendix I provides a detailed explanation of how each matching method was chosen and includes visual representations in the form of Love Plots. These plots, in combination with the balance statistics displayed in Appendix J, demonstrate that each of the three groups have excellent covariate balance.

Table 3 shows the demographic, offence, sentence, risk, and complexity characteristics across the IIRMS, IRiS and Intervention groups against their respective control groups. In percentage terms, Table 3 indicates well matched treatment and control groups across variables which are known to affect reconviction and recall rates (see Chapter 2).

Outcome Measures

To accurately assess the efficacy of the interventions under review, data were requested for recalls to custody and reconvictions with a custodial sentence. These outcomes were chosen as they are considered a broad measure of release failure, encompassing significant licence

transgressions, behaviour indicative of increasing risk, and proven reoffending. As some cases experienced both recall and reconviction within the follow-up period, a decision was made to

Table 3

Sample Characteristics

	IIRMS (N=373)	Control (N=1,492)	IRiS (N=75)	Control (N=300)	Intervention (N=448)	Control (N=1,792)
Age - M (SD)	36.2 (10.9)	36.1 (11.6)	33.0 (12.0)	32.9 (11.2)	35.7 (11.1)	34.9 (10.8)
Ethnicity - N (%)						
White	322 (86.3)	1277 (85.6)	50 (66.7)	204 (68.0)	372 (83.0)	1438 (80.2)
Non-White	51 (13.7)	215 (14.4)	22 (29.3)	88 (29.3)	73 (16.3)	350 (19.5)
Refusal	0 (0)	0 (0)	3 (4.0)	8 (2.7)	3 (0.7)	4 (0.2)
Release Year - N (%)						
2014	5 (1.3)	63 (4.2)	6 (8.0)	19 (6.3)	11 (2.5)	56 (3.1)
2015	28 (7.5)	135 (9.0)	7 (9.3)	31 (10.3)	35 (7.8)	152 (8.5)
2016	68 (18.2)	256 (17.2)	16 (21.3)	54 (18.0)	84 (18.8)	302 (16.9)
2017	115 (30.8)	357 (23.9)	20 (26.7)	92 (30.7)	135 (30.1)	493 (27.5)
2018	122 (32.7)	527 (35.3)	23 (30.7)	87 (29.0)	145 (32.4)	654 (36.5)
2019	35 (9.4)	154 (10.3)	3 (4.0)	17 (5.7)	38 (8.5)	135 (7.5)
Offence - N (%)						
Violent	190 (50.9)	736 (49.3)	43 (57.3)	174 (58.0)	233 (52.0)	937 (52.3)
Sexual	47 (12.6)	208 (13.9)	8 (10.7)	32 (10.7)	55 (12.3)	248 (13.8)
Other	136 (36.5)	548 (36.7)	24 (32.0)	94 (31.3)	160 (35.7)	607 (33.9)
Sentence Type - N (%)						
Determinate	257 (68.9)	1083 (72.6)	72 (96.0)	290 (96.7)	329 (73.4)	1326 (74.0)
Indeterminate	116 (31.1)	409 (27.4)	3 (4.0)	10 (3.3)	119 (26.6)	466 (26.0)
Sentence Length - M (SD)*	52.6 (42.3)	55.6 (47.1)	47.6 (46.1)	46.5 (43.5)	51.5 (43.2)	46.9 (39.5)
OGRS1 Score - M (SD)	44.8 (21.3)	44.5 (21.4)	48.8 (22.2)	48.0 (20.9)	45.5 (21.5)	44.2 (20.8)
OGRS2 Score - M (SD)	59.2 (21.8)	58.9 (22.4)	62.8 (22.7)	62.4 (21.5)	59.8 (22.0)	58.7 (21.9)
RoSH - N (%)						
Low	0 (0)	0 (0)	1 (1.3)	5 (1.7)	1 (0.2)	5 (0.3)
Medium	43 (11.5)	186 (12.5)	10 (13.3)	43 (14.3)	53 (11.8)	239 (13.3)
High	313 (83.9)	1244 (83.4)	61 (81.3)	242 (80.7)	374 (83.5)	1495 (83.4)
Very High	17 (4.6)	62 (4.2)	3 (4.0)	10 (3.3)	20 (4.5)	53 (3.0)
Complexity Score - M (SD)	2.8 (1.4)	2.8 (1.3)	3.0 (1.3)	2.9 (1.2)	2.8 (1.4)	2.9 (1.2)

* Determinate sentenced offenders only (sentence length for indeterminate offenders is listed as 999 years).

use the first event only. This avoids the issue of double counting, as it is often the case that a recall and reconviction for the same individual are due to the same incident. A follow-up period of 12 months was used, as this was considered a reasonable amount of time in which to observe recalls or reconvictions and is the same follow-up period as numerous other studies using similar outcome measures (e.g., Duwe & McNeeley, 2021; Shaul et al., 2016). Despite a longer follow-up period being considered, the potential impact of the Coronavirus pandemic meant that this was not possible. The final release from custody date included in the study was 31st

March 2019, meaning that the follow-up period finished on 31st March 2020, which was almost exactly when the effects of the pandemic began to be felt in the UK. Around this time, lockdown periods were enforced, and most intervention services moved to virtual working, both of which had a detrimental effect on released offenders in the community, thus having the potential to distort the findings.

The current study sought to avoid the dichotomous conclusions drawn by most previous recidivism research, which attributes *failure* to any case that is recalled or reconvicted and *success* to any case that avoids recall or reconviction during the follow-up period. The literature recognises the lack of sophistication in such research and highlights the additional value in understanding for how long a case remains incident-free (Peters et al., 2015). Anecdotally, it is accepted that there may be a significant difference between a case that is recalled or reconvicted three days after release and a case recalled or reconvicted 360 days after release. It is suggested that incarceration can have a criminogenic effect on people (Engelen et al., 2016) and that extended time in the community can increase employment opportunities, strengthen family and social support, and create the psychological conditions required for hope (Abramovaite et al., 2019). More time spent in the community, and therefore less in custody, also reduces the financial burden on the state (Ministry of Justice, 2012 & 2019).

Survival Analysis

Survival analysis examines whether an intervention influences survival rates when compared with a different intervention or no intervention (control). Survival analysis measures the time to a specific event and is particularly helpful when comparing the experiences of different groups of individuals (Flynn, 2012). In medical research, the event is often death or the progression of a disease (e.g., Austin & Schuster, 2016; Gayat et al., 2012). In forensic

research, survival analysis is becoming increasingly used to measure time before a return to a secure environment (e.g., Harrison & Alves-Costa, 2022; Penney et al., 2018). In the current study, the main outcome data, or dependent variables, were recall and reconviction. These were examined separately and as a combined measure of release failure, to determine whether a particular intervention plays a role in any of these outcomes. Survival analysis has two potential outcomes for each case. In the current study, a case either experienced the outcome under review (recall or reconviction) or they did not. Cases that did not experience the outcome of interest within the follow-up period are considered *censored* cases, as they did not contribute to the overall proportion of cases that *failed* during the follow-up period. In the current study, most censored cases simply were not recalled or reconvicted, although consideration was given to other reasons for cases being censored. Some cases could have been lost to follow-up, some may have been recalled or reconvicted but after the follow-up period and others may have passed away. As only cases with a minimum of 12 months left to serve on their Probation licence were included in the current study, no cases were lost to follow-up. It is accepted that a small number of cases may have died during the follow-up period, although there is no reason to suspect that rates of death would substantially differ between groups. It is also acknowledged that some cases across all groups will have been recalled or reconvicted after the follow-up period. For this reason, care is taken not to generalise results beyond a 12-month follow-up period.

Although detailed *life tables* describe the survival rate of a single population, with large data sets and multiple conditions, they become very difficult to interpret (Flynn, 2012). For clarity, Kaplan-Meier survival plots are used in the current study, as they provide a simple visual method for comparing the survival rates of two or more conditions over time. Whilst general patterns can be observed, survival curves are not able to tell us whether an apparent effect is statistically significant. For this level of detail, a log-rank test is required, which tests the null

hypothesis that there is no difference in survival rates between intervention and control groups. Whilst this is reported in the current study, a log rank test remains susceptible to confounding and bias, particularly when there is imbalance between groups. Whilst every effort has been made to balance covariates across groups, it is likely that some imbalance will remain. To minimise the potential for confounding and bias, a multivariate survival model is also used. Cox's proportional hazard model measures the likelihood of the hazard, in this case, recall or reconviction, occurring at any one time. It assumes that, if the likelihood of a hazard occurring in one group is double that of another, then it will remain double at any other time (Bewick et al., 2004). It is essential that this assumption is met and reported to generate reliable results (Batson et al., 2016; Chai-Adisaksopha et al., 2016). In the current study, this assumption is tested using two methods, both of which are explained below. The multivariate regression model allowed an assessment of whether individual covariates, such as age, ethnicity, offence type, risk level, and personality complexity, effect survival rates in the community. It should be noted that the regression model identified predictor variables across the whole population of a sample, meaning that the IIRMS sample and its associated control group were combined for this level of analysis. The IRiS and Intervention groups were also combined with their associated control groups for multivariate analysis. However, as group membership (control / intervention) was included as a variable, it is possible to assess whether the presence of an intervention predicted survival outcome.

Results

Overview of Outcomes

Table 4 shows the number of cases returned to custody within the follow-up period (12 months) against the number of cases still surviving in the community for each of the three intervention

Table 4

Return to Custody and Survival Data by Group Across Each Variable

	IIRMS (N=373)		Control (N=1,492)		IRiS (N=75)		Control (N=300)		Intervention (N=448)		Control (N=1,792)	
	Returned to Custody N=153 (41.0%)	Survived in the Community N=220 (59.0%)	Returned to Custody N=546 (36.6%)	Survived in the Community N=946 (63.4%)	Returned to Custody N=29 (38.7%)	Survived in the Community N=46 (61.3%)	Returned to Custody N=83 (27.7%)	Survived in the Community N=217 (72.3%)	Returned to Custody N=182 (40.6%)	Survived in the Community N=266 (59.4%)	Returned to Custody N=629 (35.1%)	Survived in the Community N=1163 (64.9%)
Age - M (SD)	33.4 (9.2)	38.3 (11.5)	34.5 (9.9)	37.0 (12.3)	33.1 (13.0)	33.0 (11.4)	31.3 (8.9)	33.5 (12.0)	33.3 (9.9)	37.3 (11.6)	33.5 (9.2)	35.7 (11.5)
Ethnicity - N (%)												
White	134 (46.1)	188 (58.4)	478 (37.4)	799 (62.6)	20 (40.0)	30 (60.0)	63 (30.9)	141 (69.1)	154 (41.4)	218 (58.6)	519 (36.1)	919 (63.9)
Non-White	19 (37.3)	32 (62.7)	68 (31.6)	147 (68.4)	8 (36.4)	14 (63.6)	18 (20.5)	70 (79.5)	27 (37.0)	46 (63.0)	109 (31.1)	241 (68.9)
Refusal	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (33.3)	2 (66.7)	2 (25.0)	6 (75.0)	1 (33.3)	2 (66.7)	1 (25.0)	3 (75.0)
Release Year - N (%)												
2014	1 (20.0)	4 (80.0)	12 (19.0)	51 (81.0)	2 (33.3)	4 (66.7)	10 (52.6)	9 (47.4)	3 (27.3)	8 (72.7)	16 (28.6)	40 (71.4)
2015	10 (35.7)	18 (64.3)	31 (23.0)	104 (77.0)	4 (57.1)	3 (42.9)	10 (32.3)	21 (67.7)	14 (40.0)	21 (60.0)	38 (25.0)	114 (75.0)
2016	30 (44.1)	38 (55.9)	67 (26.2)	189 (73.8)	5 (31.3)	11 (68.8)	7 (13.0)	47 (87.0)	35 (41.7)	49 (58.3)	68 (22.5)	234 (77.5)
2017	54 (47.0)	61 (53.0)	101 (28.3)	256 (71.7)	9 (45.0)	11 (55.0)	20 (21.7)	72 (78.3)	63 (46.7)	72 (53.3)	128 (26.0)	365 (74.0)
2018	43 (35.2)	79 (64.8)	242 (45.9)	285 (54.1)	8 (34.8)	15 (65.2)	26 (29.9)	61 (70.1)	51 (35.2)	94 (64.8)	293 (44.8)	361 (55.2)
2019	15 (42.9)	20 (57.1)	93 (60.4)	61 (39.6)	1 (33.3)	2 (66.7)	10 (58.8)	7 (41.2)	16 (42.1)	22 (57.9)	86 (63.7)	49 (36.3)
Offence - N (%)												
Violent	79 (41.6)	111 (58.4)	269 (36.5)	467 (63.5)	13 (30.2)	30 (69.8)	48 (27.6)	126 (72.4)	92 (39.5)	141 (60.5)	317 (33.8)	620 (66.2)
Sexual	11 (23.4)	36 (76.6)	55 (26.4)	153 (73.6)	5 (62.5)	3 (37.5)	7 (21.9)	25 (78.1)	16 (29.1)	39 (70.9)	67 (27.0)	181 (73.0)
Other	63 (46.3)	73 (53.7)	222 (40.5)	326 (59.5)	11 (45.8)	13 (54.2)	28 (29.8)	66 (70.2)	74 (46.3)	86 (53.8)	245 (40.4)	362 (59.6)
Sentence Type - N (%)												
Determinate	117 (45.5)	140 (54.5)	426 (39.3)	657 (60.7)	29 (40.3)	43 (59.7)	81 (27.9)	209 (72.1)	146 (44.4)	183 (55.6)	499 (37.6)	827 (62.4)
Indeterminate	36 (31.0)	80 (69.0)	120 (29.3)	289 (70.7)	0 (0.0)	3 (100.0)	2 (20.0)	8 (80.0)	36 (30.3)	83 (69.7)	130 (27.9)	336 (72.1)
Sentence Length - M (SD)*	48.7 (37.1)	55.9 (46.1)	55.9 (43.2)	55.4 (49.5)	53.3 (50.4)	43.8 (43.1)	54.9 (40.0)	43.2 (44.5)	49.6 (40.0)	53.1 (45.6)	48.8 (35.7)	45.7 (41.5)
OGRS1 Score - M (SD)	53.3 (19.8)	38.9 (20.4)	51.2 (19.4)	40.7 (21.6)	50.6 (24.2)	47.8 (21.1)	54.3 (18.3)	45.6 (21.4)	52.8 (20.5)	40.4 (20.8)	51.2 (18.7)	40.3 (20.9)
OGRS2 Score - M (SD)	67.6 (18.7)	53.4 (22.0)	65.9 (18.6)	54.8 (23.4)	64.0 (23.7)	62.1 (22.2)	68.9 (17.5)	59.9 (22.4)	67.0 (19.6)	54.9 (22.2)	66.0 (18.2)	54.7 (22.7)
RoSH - N (%)												
Low	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)	1 (20.0)	4 (80.0)	0 (0.0)	1 (100.0)	1 (20.0)	4 (80.0)
Medium	7 (16.3)	36 (83.7)	54 (29.0)	132 (71.0)	2 (20.0)	8 (80.0)	5 (11.6)	38 (88.4)	9 (17.0)	44 (83.0)	49 (20.5)	190 (79.5)
High	134 (42.8)	179 (57.2)	463 (37.2)	781 (62.8)	27 (44.3)	34 (55.7)	76 (31.4)	166 (68.6)	161 (43.0)	213 (57.0)	553 (37.0)	942 (63.0)
Very High	12 (70.6)	5 (29.4)	29 (46.8)	33 (53.2)	0 (0.0)	3 (100.0)	1 (10.0)	9 (90.0)	12 (60.0)	8 (40.0)	26 (49.1)	27 (50.9)
Complexity Score - M (SD)	3.1 (1.4)	2.6 (1.4)	3.1 (1.3)	2.6 (1.3)	2.8 (1.2)	3.1 (1.3)	3.4 (1.1)	2.7 (1.2)	3.1 (1.4)	2.7 (1.4)	3.2 (1.2)	2.7 (1.2)

* Determinate sentenced offenders only (sentence length for indeterminate offenders is listed as 999 years).

groups (IIRMS, IRiS and Intervention) compared to their respective control groups for each of the key variables used in the study.

Table 5 shows the numbers and percentages of cases recalled and reconvicted, as well as the combined returned to custody rates for all three groups compared to their respective control groups. The table also shows the rate of returns to custody within the first 100 days. Log Rank test results are included in the table to determine whether any differences between the intervention and control groups were statistically significant. Results indicate no significant difference in survival distributions between any of the three experimental groups and their respective control groups for time to recall, any return to custody or any return to custody within the first 100 days. However, when examining time to reconviction, both the IIRMS and Intervention groups showed a significantly worse survival distribution than their respective control groups.

Table 5

Recall, Reconviction and Log Rank Test Results

	Sample Size	Recalled			Reconvicted			Returned to Custody			Returned to Custody within 100 Days		
		N	%	*p*	N	%	*p*	N	%	*p*	N	%	*p*
IIRMS	373	86	23.1	.765	67	18.0	.008**	153	41.0	.181	70	18.8	.427
Control	1492	366	24.5		180	12.1		546	36.6		303	20.3	
IRiS	75	19	25.3	.303	10	13.3	.074	29	38.7	.066	21	28.0	.083
Control	300	58	19.3		25	8.3		83	27.7		55	18.3	
Intervention	448	105	23.4	.515	77	17.2	.003**	182	40.6	.061	91	20.3	.831
Control	1792	426	23.8		203	11.3		629	35.1		365	20.4	

*p<0.05; **p<0.01; ***p<0.001

Kaplan-Meier Survival Plots

To examine survival rates across the follow-up period, the Kaplan-Meier method was used to create survival plots. Kaplan-Meier survival plots show the proportion of the population

surviving at any given time and display separate lines for the intervention and control groups. Figures 4, 5 and 6 show the proportion of IIRMS, IRiS and Intervention cases, alongside their respective control groups, that survived the first 12 months without recall or reconviction. Censored cases, marked on the plots at the 365-day point, indicate that all other cases did not experience recall or reconviction within the follow-up period.

Figures 4 – 6 indicate that all three intervention groups survived at marginally higher rates initially, before survival rates dropped below that of their respective control group. Because of these findings, analysis of the first 100 days was conducted, to see if there was a treatment effect during this period. Whilst the survival plots in Appendix K indicate a positive effect of treatment initially, Table 4 confirms that none of these results were statistically significant. Figures 7 and 8 display the Kaplan-Meier curves for time to reconviction for the IIRMS and Intervention groups. Both results were statistically significant.

Figure 4

Survival to Return to Custody for IIRMS Group

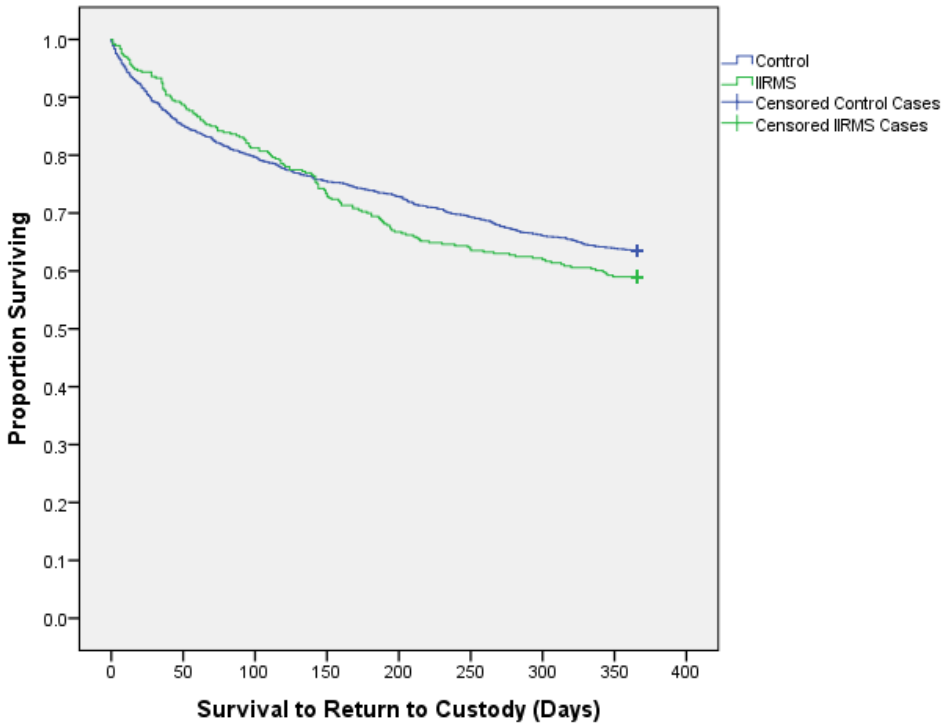


Figure 5

Survival to Return to Custody for IRiS Group

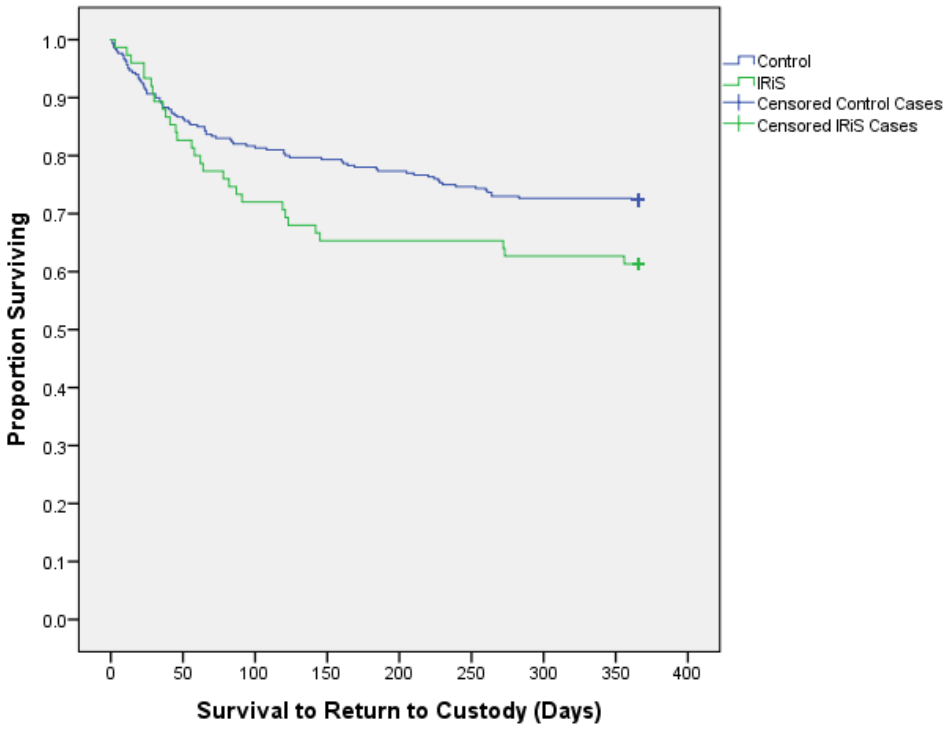


Figure 6

Survival to Return to Custody for Intervention Group

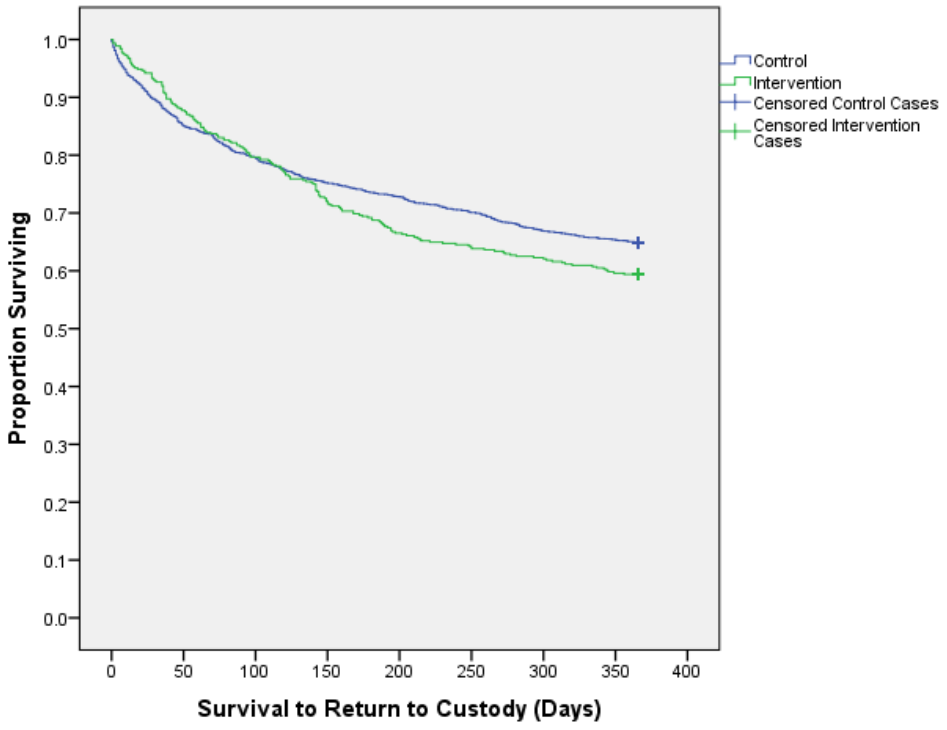


Figure 7

Survival to Reconviction for IIRMS Group

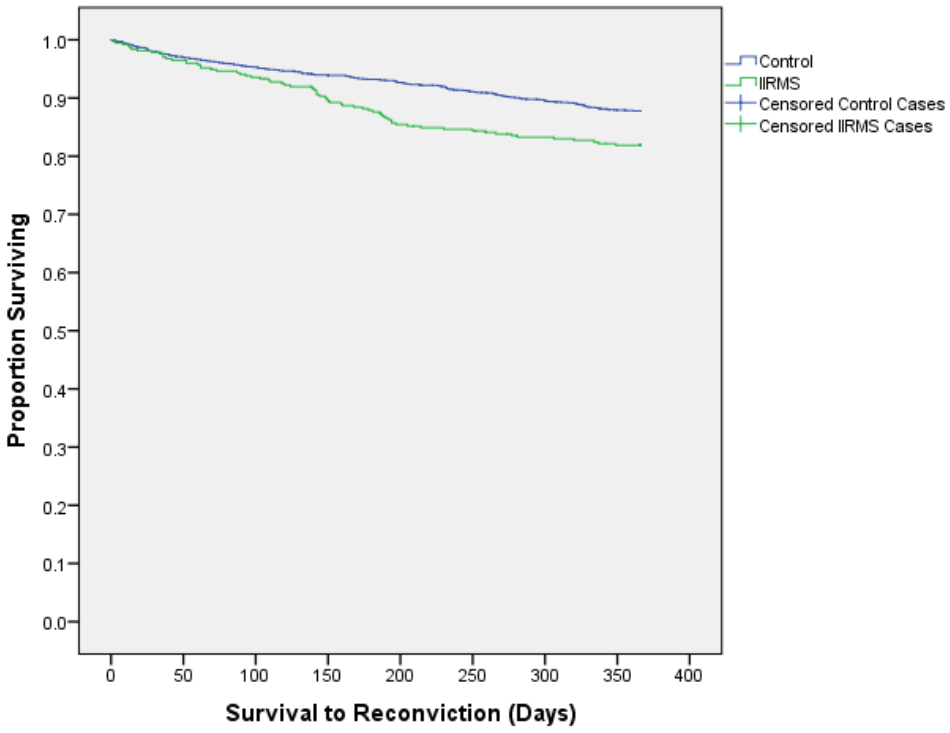
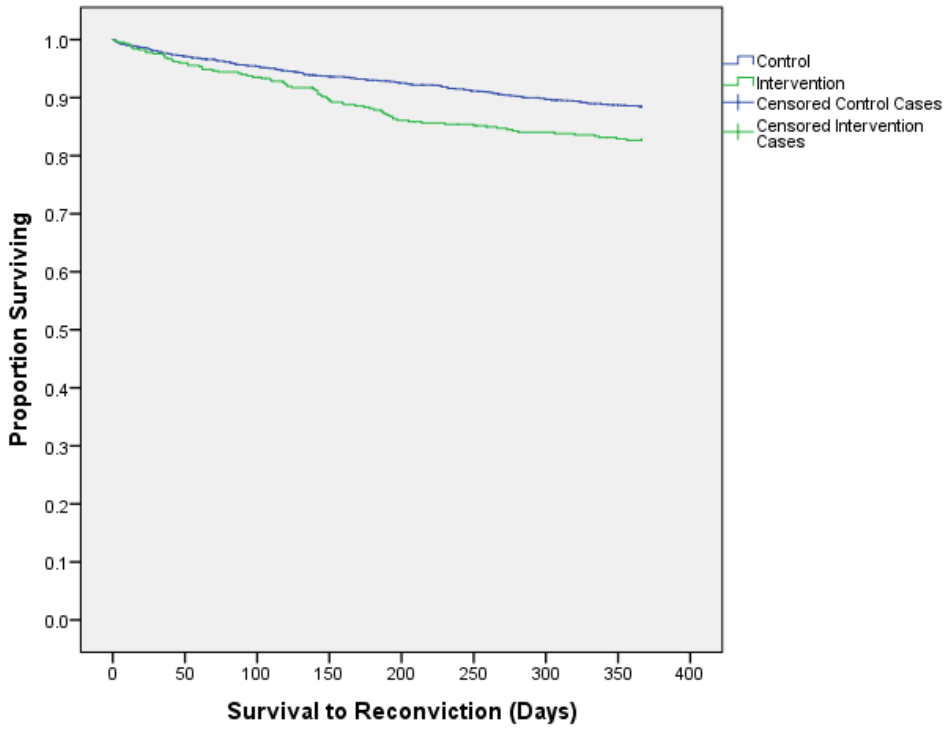


Figure 8

Survival to Reconviction for Intervention Group



Univariate Analysis

A Cox regression model was used to establish the effect of individual variables on survival in the community. The treatment and control groups were combined into single populations for this analysis and group membership was included as a covariate. These populations are referred to as the IIRMS, IRiS and Intervention populations. The variables assessed for influence over the outcome measures are shown in Table 6, below. These are the same variables as were used to match the treatment and control groups, with the addition of the individual OPDP screening items, as well as the overall complexity score. Appendix I describes how the complexity score was created. Table 6 displays the hazard rate for each covariate. The hazard rate is the risk that the event of interest, in this case return to custody, will occur within the next time period. As time is measured in days, the hazard rate can be interpreted as the risk of a return to custody within the next day, i.e., the current risk. Hazard rates greater than 1 indicate that the risk of a return to custody increases with the presence of or an increase in that variable.

Table 6 shows that within the IIRMS population, including the relevant control group, the presence of, or an increase in a unit of, all three risk variables[2] and all five OPDP screening variables, including overall complexity score[3], significantly increased the risk of a return to custody. An increase in RoSH category and the presence of OASys items, childhood difficulties and challenging behaviour each increased the chances of a return to custody by a

[2] Offender Group Reconviction Scale 1 (OGRS1; HR=1.02; 95% CI=1.02-1.03; p<.001), Offender Group Reconviction Scale 2 (OGRS2; HR=1.02; 95% CI=1.02-1.03; p<.001), Risk of Serious Harm (RoSH; HR=1.58; 95% CI=1.31-1.92; p<.001)

[3] OASys items (HR=1.62; 95% CI=1.35-1.93; p<.001), childhood difficulties (HR=1.63; 95% CI=1.37-1.93; p<.001), mental health (HR=1.16; 95% CI=1.00-1.35; p<.05), self-harm/suicide attempts (HR=1.27; 95% CI=1.09-1.47; p<.01), challenging behaviour (HR=1.58; 95% CI=1.36-1.83; p<.001), complexity score (HR=1.27; 95% CI=1.20-1.34; p<.001)

factor of 1.6. For the same population the presence of, or an increase in a unit of, age[4], ethnicity[5], sentence type[6], and sentence length[7] significantly decreased the risk of a return to custody. This indicates that with an increase in age and sentence length comes a decrease in the risk of a return to custody and that non-white and indeterminate sentenced men have a reduced risk of returning to custody. Group membership (control v IIRMS) had no significant effect on survival in the community.

Examination of the IRiS population, including the relevant control group, shows that the presence of, or an increase in a unit of, OGRS1 scores[8], OGRS2 scores[9], OASys items[10], childhood difficulties[11], challenging behaviour[12], and complexity scores[13] were significantly associated with a higher risk of returning to custody. The presence of challenging behaviour increased the likelihood of a return to custody in this cohort by a factor of 2.2. Group membership (control v IRiS) had no significant effect on survival in the community.

It is acknowledged that the IIRMS group had a significant impact on the outcome of the Intervention group, as it accounted for the majority of the population, although it should also be noted that an independent control group was matched to the Intervention population. The results for the Intervention group were similar to that of the IIRMS group, with the exception of the OPDP screening mental health variable, which had no significant effect on the risk of

[4] HR=0.98; 95% CI=0.97-0.99; p<.001
[5] HR=0.79; 95% CI=0.63-0.99; p<.05
[6] HR=0.65; 95% CI=0.54-0.78; p<.001
[7] HR=0.99954; 95% CI=0.999-0.9997; p<.001
[8] HR=1.02; 95% CI=1.01-1.03; p<.01
[9] HR=1.02; 95% CI=1.01-1.03; p<.01
[10] HR=1.73; 95% CI=1.07-2.81; p<.05
[11] HR=1.63; 95% CI=1.06-2.50; p<.05
[12] HR=2.21; 95% CI=1.47-3.32; p<.001
[13] HR=1.27; 95% CI=1.08-1.48; p<.01

Table 6

Influence of Individual Covariates on Survival in the Community

		IIRMS (N=1,865)				IRiS (N=375)				Intervention (N=2,240)			
		Hazard Rate (HR)	95% CI for HR Lower	95% CI for HR Upper	*p*	Hazard Rate (HR)	95% CI for HR Lower	95% CI for HR Upper	*p*	Hazard Rate (HR)	95% CI for HR Lower	95% CI for HR Upper	*p*
	Group (Control / Intervention)	1.13	0.94	1.35	0.187	1.48	0.97	2.26	.068	1.17	0.99	1.38	.061
Demographic Variables	Age	0.98	0.97	0.99	**<.001*** **	0.99	0.97	1.01	.179	0.98	0.98	0.99	**<.001*** **
	Ethnicity (White / Non-White)	0.79	0.63	0.99	**.039***	0.68	0.45	1.04	.077	0.82	0.68	0.98	**.034***
Conviction Variables	Offence (Violent / Other)	0.97	0.84	1.13	.723	1.16	0.80	1.68	.432	1.08	0.94	1.24	.292
	Sentence Type (Determinate / Indeterminate)	0.65	0.54	0.78	**<.001*** **	0.50	0.12	2.02	.330	0.65	0.54	0.77	**<.001*** **
	Sentence Length	1.00†	1.00†	1.00†	**<.001*** **	1.00†	1.00†	1.00†	.656	1.00†	1.00†	1.00†	**<.001*** **
Risk Variables	OGRS1 Score	1.02	1.02	1.03	**<.001*** **	1.02	1.01	1.03	**.002****	1.02	1.02	1.03	**<.001*** **
	OGRS2 Score	1.02	1.02	1.03	**<.001*** **	1.02	1.01	1.03	**.002****	1.02	1.02	1.03	**<.001*** **
	RoSH	1.58	1.31	1.92	**<.001*** **	1.42	0.94	2.14	.094	1.88	1.57	2.26	**<.001*** **
Personality Variables	OASys Items (Absent / Present)	1.62	1.35	1.93	**<.001*** **	1.73	1.07	2.81	**.026***	1.60	1.35	1.89	**<.001*** **
	Childhood Difficulties (Absent / Present)	1.63	1.37	1.93	**<.001*** **	1.63	1.06	2.50	**.025***	1.46	1.25	1.71	**<.001*** **
	Mental Health (Absent / Present)	1.16	1.00	1.35	**.046***	0.79	0.54	1.14	.209	1.09	0.95	1.26	.202
	Self Harm / Suicide Attempts (Absent / Present)	1.27	1.09	1.47	**.002****	1.12	0.77	1.63	.543	1.18	1.03	1.35	**.020***
	Challenging Behaviour (Absent / Present)	1.58	1.36	1.83	**<.001*** **	2.21	1.47	3.32	**<.001*** **	1.66	1.44	1.91	**<.001*** **
	Complexity Score (0-5)	1.27	1.20	1.34	**<.001*** **	1.27	1.08	1.48	**.003****	1.25	1.18	1.32	**<.001*** **

*p<0.05; **p<0.01; ***p<0.001

† Unrounded figure <1

returning to custody. Again, group membership (control v Intervention) had no significant effect on survival in the community.

Multivariate Analysis

Cox proportional hazard model was used to construct a multivariate predictive model for return to custody. The model combined both the intervention and control groups for each population, and group membership (control v intervention) was included as one of the covariates. Except for group membership, only covariates showing a significant univariate effect on return to custody were included in the model. Group membership was included, despite not having a significant effect on outcome as a single variable, as the current study focuses on the effect of intervention on survival in the community. The model was constructed using a backward step-wise approach, meaning that all relevant covariates were initially entered into the model before some were removed according to the effect they had on the predictive accuracy of the model. The model was constructed in two phases with all covariates, except group membership, being added in block 1 and group membership being added in block 2. This allowed the overall effect of each intervention to be analysed when added to the model.

Cox proportional hazard model requires the proportional hazard assumption to be met. Simply put, this means that the risk of returning to custody must be the same throughout the follow-up period. There are various means of assessing this assumption and two methods are used here: visual assessment of Kaplan-Meier curves and analysis of each covariate's interaction with time. The model is only used to predict any return to custody, so only the relevant Kaplan-Meier curves were assessed for the proportional hazard assumption.

Analysis of the proportional hazard assumption, using both methods described above, is contained in Appendix L. For some variables, the assumption was not met, resulting in the

need for 'dummy' variables accounting for the interaction with time to be included in the regression models.

Table 7 displays the final regression model for the IIRMS population and includes all covariates that increased the predictive accuracy of the model. This method of analysis requires a *reference group* to be identified for each categorical variable with more than 2 categories. Reference groups were identified for RoSH and Complexity Score with the lowest categories being used for both (medium risk and 0, respectively). For these two variables, Table 7 shows the hazard rate for each category when compared to the reference group. The model indicates that being non-white, having an indeterminate sentence and having a history of mental health difficulties were protective factors for survival in the community. As a single variable, mental health difficulties significantly increased the risk of a return to custody but when accounting for the confounding effect of other covariates, this changed to a significant protective factor. OGRS2 scores[14], RoSH[15], and complexity scores[16] were all associated with a significant increase in the risk of a return to custody. Table 7 shows that being high RoSH increased the risk of a return to custody, when compared to the reference group (medium RoSH) by a factor of 2.5 (HR=2.55). Being very high RoSH increased the risk of a return to custody, compared to the reference group, by a factor of over 3 (HR=3.26). Having a complexity score of 3, 4 or 5 significantly increased the risk of a return to custody, when compared to the reference group (0), by factors of 1.8 (HR=1.76), 2.0 (HR=1.95) and 2.8 (HR=2.82), respectively. The addition of group membership had no effect on the predictive accuracy of the model.

The regression model included the significant interaction with time shown by ethnicity, sentence type, sentence length and RoSH. This shows that the effect on the hazard rate of

[14] HR=1.02, 95% CI=1.02-1.02, p<.001
[15] p<.001
[16] p<.001

sentence type and being high RoSH significantly reduced with the passing of each unit of time. This suggests that the risk of a return to custody associated with having a determinate sentence and being high RoSH is strong initially but diminishes throughout the 12-month follow-up period. Conversely, the risk of a return to custody associated with being white appears to significantly increase with the passing of each unit of time.

Table 7

Final Regression Model for IIRMS Population

	Hazard Rate	95% CI for HR		*p*
		Lower	Upper	
Ethnicity (White / Non-White)	0.627	0.431	0.913	.015*
Sentence Type (Determinate / Indeterminate)	0.544	0.401	0.738	<.001***
OGRS2 Score	1.020	1.016	1.024	<.001***
RoSH (Med - reference group)				<.001***
RoSH (High)	2.545	1.581	4.098	<.001***
RoSH (Very High)	3.259	1.711	6.209	<.001***
Complexity Score (0 - reference group)				<.001***
Complexity Score (1)	0.988	0.559	1.746	.966
Complexity Score (2)	1.364	0.799	2.329	.256
Complexity Score (3)	1.764	1.029	3.024	.039*
Complexity Score (4)	1.95	1.114	3.414	.019*
Complexity Score (5)	2.819	1.584	5.019	<.001***
Mental Health (Absent / Present)	0.783	0.641	0.955	.016*
Ethnicity * Time	1.002	1.000	1.005	.021*
Sentence Type * Time	0.983	0.972	0.994	.003**
Sentence Length * Time	1.000	1.000	1.000	<.001***
RoSH (Med) * Time (reference group)				.008**
RoSH (High) * Time	0.996	0.994	0.999	.002**
RoSH (Very High) * Time	0.997	0.993	1.001	.150
Group (Control / Intervention)	1.101	0.918	1.32	.300

*p<0.05; **p<0.01; ***p<0.001

Regression model contains 10 steps

Table 8 shows the final regression model for the IRiS population. Four variables were included in the final model and none of the variables demonstrated a significant interaction with time. Elevated OGRS1 scores[17] and the presence of challenging behaviour[18] were both associated with a significant increase in the risk of a return to custody for men in the IRiS population. The addition of group membership had no effect on the predictive accuracy of the model.

Table 8

Final Regression Model for IRiS Population

	Hazard Rate	95% CI for HR		*p*
		Lower	Upper	
OGRS1 Score	1.010	1.000	1.020	.046
OASys Items (Absent / Present)	1.577	0.966	2.575	.069
Challenging Behaviour (Absent / Present)	1.936	1.274	2.944	.002
Group (Control / Intervention)	1.325	0.865	2.030	.196

*p<0.05; **p<0.01; ***p<0.001

Regression model contains 4 steps

Table 9 displays the final regression model for the combined Intervention population. Higher sentence length[19], OGRS2 scores[20] and RoSH categories[21], as well as the presence of a history of self-harm/suicide attempts[22] and the presence of challenging behaviour[23] all significantly raised the risk of a return to custody for the Intervention population. Having an indeterminate

[17] HR=1.01, 95% CI=1.00-1.02, p<.05
[18] HR=1.94, 95% CI=1.27-2.94, p<.01
[19] HR=1.00, 95% CI=1.00-1.01, p<.001
[20] HR=1.02, 95% CI=1.02-1.03, p<.001
[21] p<.001
[22] HR=1.15, 95% CI=1.00-1.33, p<.05
[23] HR=1.33, 95% CI=1.15-1.54, p<.001

sentence[24] significantly reduced the likelihood of a return to custody. Again, the addition of group membership had no effect on the predictive accuracy of the model.

The model included two covariates that demonstrated an interaction with time. It is not possible from the data to determine the direction of the interaction with time for sentence length, suggesting that the effect is small. The effect on return to custody of the presence of childhood difficulties significantly increased with the passing of each unit of time. However, childhood difficulties as a covariate was not included in the final model.

Table 9

Final Regression Model for Intervention Population

	Hazard Rate	95% CI for HR		*p*
		Lower	Upper	
Sentence Type (Determinate / Indeterminate)	0.009	0.001	0.058	<.001***
Sentence Length	1.004	1.002	1.006	<.001***
OGRS2 Score	1.021	1.016	1.025	<.001***
RoSH (Low - reference group)				<.001***
RoSH (Medium)	0.998	0.138	7.229	.999
RoSH (High)	1.683	0.236	12.008	.603
RoSH (Very High)	2.399	0.329	17.52	.388
OASys Items (Absent / Present)	1.186	1.000	1.407	.050
Self Harm / Suicide Attempts (Absent / Present)	1.153	1.001	1.329	.049*
Challenging Behaviour (Absent / Present)	1.327	1.148	1.535	<.001***
Sentence Length * Time	1.000	1.000	1.000	<.001***
Childhood Difficulties * Time	1.002	1.001	1.003	.001**
Group (Control / Intervention)	1.109	0.939	1.310	.224

*p<0.05; **p<0.01; ***p<0.001

Regression model contains 7 steps

[24] HR=0.01, 95% CI=0.00-0.06, p<.001

Discussion

Summary of Main Findings

This is the first large-scale, long-term examination of IIRMS and IRiS using robustly matched control groups. Analysis was unable to detect any effects of either intervention on survival in the community, although it was established that IIRMS cases were reconvicted at a significantly higher rate than the control group. It is acknowledged that the results of the Intervention group, a combination of the IIRMS and IRiS populations, were significantly influenced by the IIRMS population, who formed the majority of this cohort. For this reason, the results of the Intervention population are discussed below, but given less emphasis than the IIRMS and IRiS results.

Examining covariates individually revealed that strong predictors of shorter survival times for the IIRMS population were being young, white, having a determinate sentence and having a shorter sentence length. As was expected, having higher risk scores on OGRS1, OGRS2 and RoSH were also strong predictors of shorter survival times within the 12-month follow up period. All five covariates from the OPDP screening process were significant predictors of shorter survival times, with the presence of OASys items, childhood difficulties and challenging behaviour showing the strongest effects. Demographic and conviction covariates were not predictive of survival times for the IRiS population, although high OGRS1 and OGRS2 scores were. Perhaps due to a smaller sample size, it was only possible to detect a negative effect on survival time due to the presence of OASys items, childhood difficulties and challenging behaviour for the IRiS population. For the Intervention population, being young, white, having a determinate sentence and a shorter sentence length were predictive of a shorter survival time in the community. Higher risk scores on OGRS1, OGRS2 and RoSH were also predictive of shorter survival times. Mental health was the only OPDP screening variable

whose presence did not have a significant relationship with shorter survival times in the community.

Accounting for the confounding effect of covariates on each other, multivariate analysis demonstrated that being white, having a determinate sentence, higher OGRS2 scores and a higher RoSH rating were associated with a shorter survival time in the community for the IIRMS population. Complexity scores of three and above, as well as the absence of mental health difficulties were also predictive of shorter survival times. For the IRiS population, higher OGRS1 scores and the presence of challenging behaviour were both predictive of shorter survival times in the community. For the Intervention population, having a determinate sentence, a longer sentence length, a higher OGRS2 score and a higher RoSH rating are associated with shorter survival times in the community. The presence of a history of self-harm / suicide attempts and challenging behaviour were also predictive of shorter survival times for the Intervention population. Group membership did not influence survival time in the community for any of the three populations.

The fact that there was no discernible effect of either IIRMS or IRiS on survival time in the community raises questions about the efficacy of such interventions. It was expected that an intervention effect would be shown and that those cases managed through either of these interventions would survive longer in the community. Whilst this effect was not identified, it was noticeable that the cohort offered an intervention were returned to custody at a lower rate immediately after they started in service or were released from custody, when compared to the cohort managed by Probation as Usual (control). Whilst none of these results were statistically significant, it may be that the structure and support of an intervention assisted some men during this initial period but was not sufficient to have a longer-term effect. After this initial period, however, the survival rate across all three groups fell below that of the control groups. Whist

this was not a statistically significant effect, concerns are raised that offering an intervention may result in a lower survival rate than managing men via a standard Probation licence.

However, there may also be an issue of selection bias at the point that cases are chosen for either of the specific interventions under review. Cases are often referred to IIRMS and IRiS when there is evidence of an increase in risk, a decrease in wellbeing or recent instability, often connected to substance misuse, housing or relationship status. Whilst these individuals may be comparable to their control group according to the relatively static variables used in the current study, they may be incomparable in terms of recent stability. The only robust research method of accounting for these differences is likely to be an RCT design, which come with some ethical and practical concerns when working with high-risk forensic clients. This is explored further below.

Current Findings in the Context of Similar Studies

The current study does not replicate the findings of a small cohort of literature that has established a treatment effect for community-based interventions. Clark (2015) demonstrated that the high-risk revocation reduction (HRRR) programme was able to significantly reduce recalls and reconvictions. Braga et al. (2009) established that the Boston Reentry Initiative was significantly associated with a reduction in both overall and violent arrest rates. Antonio and Crossett (2017) report a significant reduction in recidivism for released offenders completing a cognitive life skills programme when compared to a matched control group. This effect was even stronger for the high-risk group, however, the same study established that the treatment group were reconvicted at a significantly higher rate, matching the findings of the current study.

The current results are consistent with several other studies assessing the impact of interventions on recidivism, re-arrest and returns to custody. While the current study was able to detect marginally better outcomes for intervention cases during the initial period in service, there was no statistically significant difference in survival rates between intervention and control groups across the 12-month follow-up period. Lattimore and Visher (2009) found similar results with marginally improved, but non-significant, outcomes regarding housing, employment, substance misuse, and self-reported criminality, but no significant differences between intervention and control groups regarding recidivism measures, for both adult and juvenile cohorts. Nathan et al. (2019) also found no significant effect of treatment using an RCT design and an intent-to-treat approach, with higher levels of detected offending in the treatment group. Although focusing specifically on substance misusing repeat offenders, Shaul et al. (2016) found no effect of intervention when using recidivism and time to reoffend measures across a 12-month follow up period. Regarding revocation of parole (recall), Bouffard and Bergeron (2006) detected no difference between treatment and control groups when examining a Serious and Violent Offender Reentry Initiative (SVORI). The current study is also consistent with Hanson et al.'s (2004) findings that men convicted of a sexual offence offered community treatment and followed for up to 12 years, showed no measurable reduction in recidivism compared to a matched control group. The consistent inability of studies to detect treatment effects for community-based interventions requires further exploration.

Return to Custody Findings

Higher reconviction rates for men offered an intervention are concerning but not surprising, and several previous studies have established a similar pattern (e.g., Nathan et al., 2019). Whilst care was taken to ensure key variables were equally matched across treatment and

control groups, the choice to offer an intervention is typically based on an individual's risk, wellbeing, or stability at a particular time. Men offered an intervention by IIRMS or IRiS are likely to have experienced a recent downturn in stability or wellbeing or a recent upturn in perceived risk, thus attracting the attention of such services. Whilst their long-term offending profile, risk level or complexity score might match that of others in the control group, it is expected that their 'real time' risk at the point of selection might be significantly higher than that of their matched control group counterparts. This is a significant drawback of an observational cohort design, as it may be the selection of an individual into a treatment group than introduces bias into the process. An RCT design in which all cases are selected for treatment and then randomly allocated to either a treatment or control group would avoid this issue, despite raising the ethical concerns discussed earlier.

Men released into the community and offered IIRMS or IRiS benefit from a range of different opportunities not afforded to most people managed on a Probation licence. This generally results in higher levels of contact with professionals in both OPDP services, Probation teams and the Police, among others. This increased level of contact equates to an increase in monitoring, making the detection of any further offending or risk related behaviour more likely. This may particularly be the case for men in IRiS, as this is an Integrated Offender Management (IOM) led project requiring regular contact with the Police. It should also be considered that reconvicting someone already on Probation licence is a Crown Prosecution Service (CPS) decision, influenced by many factors. It is often the case that men recalled on licence are not reconvicted for a further offence, as it may not be in the public interest. These decisions are likely to be affected by the seriousness of the offence and the availability of evidence, which may differ between intervention and control cases, due to the increased monitoring of intervention cohorts. Further research examining reconviction decisions for men on Probation licence would significantly improve our understanding of these findings. It would also be

helpful to examine self-reported reoffending rates, such as those used by Lattimore and Visher (2009), as this would remove the issues caused by anomalies in the court process.

Questions are also raised about the specific aims of both IIRMS and IRiS and how they have sought to meet these aims. Both services state their intention to reduce reoffending, but there are a range of methods of achieving this aim. Whilst reducing recall to custody is a stated objective for the Probation Service, it can also be a necessary tool to prevent the risk of further offending. With additional monitoring of intervention cases, it should be considered that this increases the available information on which recall decisions are made. Whilst it is obviously not possible to measure the effect of recall decisions on the avoidance of further offences in the community, this should be considered when interpreting the current findings.

Whilst there is a high degree of confidence in the accuracy of the recall and reconviction data used in the current study, there are several subtleties within the data that it has not been possible to examine. The IIRMS cohort were recalled at a marginally lower rate than their control counterparts and the IRiS cases were recalled at a higher rate than their control counterparts. Whilst neither of these results represented a significant difference in recall rates, it was not possible to accurately determine the reason for recall for either cohort. Some cases are recalled for technical breaches of licence conditions that don't necessarily indicate an increase in risk of reoffending. Other cases are recalled for suspected offences that do not meet the threshold for prosecution. For example, Ryan et al (2022) established that, of those recalled whilst supported by the Resettle IIRMS, *risky behaviour* was given as the reason in 39% of cases. This highlights the need for further research examining reasons for recall and whether interventions status plays a role in any differences found.

Given that IIRMS cases are likely to be referred at a time when there are increased concerns about wellbeing and risk, the finding that IIRMS cases survive at a marginally higher rate

during the first 100 days is noteworthy. The intervention model developed for IIRMS is based on a relational model of care (HMPPS & NHS, 2021), which requires the development of a therapeutic alliance between the service user and key staff members. As this takes time and as many of this cohort struggle to trust professionals in positions of authority, it is possible that there is another function of this relative success. The simple offer of a structured service, in a time of need, may be enough to provide some initial hope and containment for individuals. This is consistent with Ryan et al.'s (2019) qualitative research findings, which highlight 'taking a chance' as a reason for engaging in IIRMS. Further research, particularly of a qualitative nature, will clarify the specific mechanisms of success during this early period, as well as understanding how this can be translated into longer-term success.

Regarding the drop off in efficacy after an initial period within IIRMS and IRiS, there are several factors to consider. Previous research (e.g., Braga et al., 2009; Clark, 2015) has suggested that successful community-based interventions have been multimodal, offering employment assistance, mentoring, and monitoring, as well as psychological and behavioural treatment. Whilst both IIRMS and IRiS can offer support in these areas, it is not possible to determine the exact service that was offered to each person, making it challenging to conclude what worked or what was missing. The current findings, along with previous research, indicate that intervention services are likely to be more successful with a multimodal approach, focusing on lifestyle, addiction, wellbeing, and risk areas.

It has not been possible from the current data to fully understand the level of engagement with treatment shown by the intervention cohorts. Inclusion criteria stipulated that men were offered and started in a service but made no assumption about the level or longevity of engagement. Data on when men left a service was not available, meaning that cases could have started in a service, left the service, and been recalled or reconvicted subsequently, telling us little about the efficacy of each intervention. As is described in the methodology section above, pre- and

post-evaluation measures were requested but were not available with the required consistency or accuracy. These measures would have enabled an approximation of engagement, at least in terms of longevity, which would have led to a better understanding of the effect of engagement on outcomes. This is particularly relevant to community cases, where it has been found that non-completion of treatment can elevate levels of risk when compared with those offered no treatment (McMurran & Theodosi, 2007).

Consideration also needs to be given to the control group, who were not offered support through IIRMS or IRiS. Whilst care was taken to select the control group from an area where there was no comparable service in operation, they could have been offered employment support, mentoring, substance misuse support or other services that may have affected their survival chances. Further research should attempt to control for levels of engagement in services, as well as for the types of support that comparison groups may be offered. This is best done using an RCT design or by using smaller sample sizes and a qualitative research design.

Predictor Variables

Independent of whether an intervention was offered, being white was significantly associated with shorter survival times in the community for both the combined IIRMS and Intervention populations. Ethnicity was also included in the multivariate predictor model for the IIRMS population, suggesting that being white had a negative effect on survival, independent of other variables. Whilst it should be acknowledged that over 80% of the IIRMS and Intervention populations were white, this is broadly consistent with recent data on ethnic diversity among probationers in the UK, which found that 84% were white (MoJ, 2021). As these finding relate to the combined intervention and control groups, it is not possible to comment on whether

intervention services are providing equitable support to white and non-white service users. In order to better understand these data, a detailed examination of outcomes for different ethnic groups is required, something that is beyond the scope of the current study. This would also have to include an analysis of the offences committed to see if there is a correlation between ethnicity and recidivism, offence type and recidivism, or both. Qualitative research on the experiences of white and non-white intervention completers would also be of assistance in understanding these data.

Regardless of whether an intervention was offered, higher OGRS and RoSH scores were significant predictors of shorter survival times for intervention groups generally, although RoSH was not a predictor for the IRiS population. Whilst this is not surprising, it supports the strength of these assessments in understanding those cases likely to struggle most in the community. Similarly, the presence of OASys items related to anti-social traits, childhood difficulties and challenging behaviour were predictive of shorter survival rates in the community, indicating that men with these experiences and behaviours are likely to struggle most in the community. These findings are also supportive of the OPDP screening items' ability to identify complex cases that require additional support on release from custody. It is suggested that risk and OPDP variables could be used together to better understand those men who may have a higher need for support once released from custody.

For the combined IIRMS and control group, as well as the combined Intervention and control group, it was established that shorter sentence lengths and determinate sentences were significantly associated with shorter survival times. This was an expected finding for several reasons. Sentence length is a proxy for the severity of offence, indicating that those with shorter sentences are likely to have been convicted of less serious offences. Such offences often come with no mandated custodial treatment and shorter periods in custody mean that there is less time to complete any treatment that is mandated. In addition, the deterrent effect of prison for

those serving shorter sentences is less than for those serving lengthy sentences. People being managed on shorter sentences are also typically housed on more transient and chaotic wings, resulting in limited stability. People on shorter sentences are also less likely to be released to a supportive hostel environment and perhaps more likely to experience instability. Several of these points are also relevant to those convicted of determinate sentences, with such offenders given fewer treatment opportunities and perhaps experiencing a weaker deterrent effect. In addition, people released on indeterminate sentences can be recalled to custody for many years after their release and, in the case of life sentenced prisoners, for the remainder of their lifespan. Collectively, these factors are likely to have a significant deleterious effect on survival chances for those convicted of shorter determinate sentences.

Cox regression models identified variables that were predictive of outcome once the confounding effect of other covariates were controlled for. As expected, several risk related variables were predictive of lower survival rates in the community across all three populations. Complexity scores of three and above were also predictive of shorter survival times for the IIRMS population. Specifically, the presence of challenging behaviour in the IRiS population led to an almost two-fold increase in the risk of returning to custody. As this finding relates to the combined IRiS and control groups, it cannot be explained entirely by the IRiS group having greater contact with the Police. Similarly, in the Intervention population both the presence of challenging behaviour and a history of self-harm or suicide attempts were predictive of lower survival rates in the community. Again, this applied to men in the intervention and control groups. These findings offer further support for the use of risk and complexity indicators to identify those cases that are likely to require additional support in the community.

Perhaps counter intuitively, the presence of mental health issues within the IIRMS population was a protective factor, ensuring longer survival times for the combined IIRMS and control groups. It is postulated that men with a history of mental health issues may be more likely to

receive additional support from other services in the community, perhaps lowering their risk of returning to custody, which was not captured in the current study. The presence of mental health issues may also affect the interpretation of risk by professionals, potentially increasing compassion or making the link between behaviour and risk harder to determine, therefore creating a higher recall threshold for professionals. It should also be considered that men with a history of mental health difficulties may be less inclined or less able to be transparent with professionals about risk related behaviours. An exploration of this sub-population of OPDP cases would contribute significantly to the current literature, as it would allow for a better understanding of the type of support offered and the personality traits displayed that appear to be protective with regard to returning to custody within the first 12 months after release.

Sentencing variables were included in two of the final regression models, for the IIRMS and Intervention populations, suggesting that they may help to clarify which cases are at higher risk of returning to custody within 12 months. For both populations, being an indeterminate sentenced offender increased survival time in the community. Whilst indeterminate sentences are typically given for more serious or concerning offences, they also tend to be associated with higher levels of treatment in custody and a parole panel is also required to assess risk as manageable in the community prior to release. In addition, men with indeterminate sentences may feel that they have more to lose should they be returned to custody, as they are likely to spend longer in prison than their determinate sentenced counterparts. For the Intervention population, sentence length had a significant effect on outcome, with longer sentences decreasing survival time in the community, once controlling for the confounding effect of other covariates. Sentence length may be a proxy measure of institutionalisation, potentially explaining why men released from longer sentences are returned to custody at a faster rate than those on shorter sentences. Age was a univariate predictor of survival in the community, with a reduction in risk of returning to custody with each unit increase in age. This finding is similar

to that of Ryan et al (2022), who established that older offenders received fewer reconvictions within their cohort of 45 released offenders supported by the Resettle IIRMS. However, once the confounding effects of other covariates were accounted for, age was not included in any of the final regression models in the current study. A possible explanation is that sentence length and age share a significant proportion of variance, as a longer sentence length will inevitably result in a higher age on release.

Strengths and Limitations

The current study had access to a fully populated dataset of over 4,000 cases across two intervention groups and a control sample. This enabled a thorough and statistically robust matching process, using ten highly relevant covariates. Although the matching process inevitably discarded cases, the groups remained large resulting in a good Effective Sample Size able to detect significant results within the data. The covariates used in the matching process were identified through a combination of a thorough review of relevant literature and expertise within the OPDP community. However, it is possible that variables not included in the matching process also influenced outcomes, thus limiting the balance between intervention and control groups. The balance data across the variables used indicated well matched intervention and control groups across all three samples. However, it is recognised that the complexity score relies on a dichotomous assessment of the presence or absence of five indicators. This does not allow for a deeper understanding of the strength of each variable, meaning that someone with relatively minor mental health issues may be matched with someone with a serious mental illness. However, the rigorous process used to identify relevant covariates and the techniques used to match groups approximates a randomised control trial and minimised the risk of significant results being due to population differences across samples.

Survival analysis is reliant on accurate *time to event* data. The current study used return to custody data, which is reliably recorded, as Probation Practitioners are required to record recalls and reconvictions when they occur. Whilst overall return to custody data is reliable, there is less certainty regarding the reason for reincarceration. Across the total population included in the study, 23.6% were recalled and 12.5% were reconvicted. There are a range of reasons why this may not be an accurate reflection of reoffending, for example it may not always be in the public interest to pursue a conviction against someone who has already been recalled to custody. The data is also subject to variations in thresholds for recall, meaning that returns to custody can be influenced by practitioners, as well as people on probation. Although care was taken to ensure the correct risk and complexity indicators were collected for the period under review, it is recognised that this is only possible if accurate information is entered onto the two data recording systems used. Whilst there was no missing data, it is likely that some of the risk and complexity information was not accurate for some cases. This may particularly be the case for the OPDP screening information used to create the complexity score. Since the start of the study period in April 2014, significant training and awareness work has been carried out to increase both compliance with and accuracy of the OPDP screening tool. Whilst there is no reason to believe this has affected intervention and control groups disproportionately, this should be considered when accounting for the results.

As survival analysis is time dependent, it requires a start date for each case. For control cases, the date each case was released from custody was used as the first day they were at risk of recall or reconviction. For intervention cases, the date they were accepted into either IIRMS or IRiS was used as the first day they were at risk of recall or reconviction. If this date was before release, the subsequent release from custody date was used instead. For many intervention cases, the start date was the same as the release from custody date. However, it is recognised that for other intervention cases, their start in service date was after their release

date, sometimes significantly so. This means that the follow up period for many intervention cases did not include the initial period of release, which we know to be one of the hardest transition periods for high-risk men (Polaschek et al., 2018; Nadesu 2007). Whilst this may suggest that the results are biased in favour of a treatment effect, another factor requires consideration: cases referred to IIRMS or IRiS following a period in the community may have been referred in response to a deterioration in wellbeing or an increase in risk related behaviour. Thus, their initial period in service may be as volatile as the initial release period and sometimes more so. It is not possible to determine the extent to which this affected the current study.

Conclusions and Recommendations for Further Research

Whilst it is disappointing that an intervention effect was not able to be established with regard to survival time in the community, a number of important discussion points are raised. These include consideration of the population being targeted for community-based interventions, as well as the type of services offered to them. It is noted that selection into a service may introduce an element of bias itself, as selection is typically based on increasing concerns around wellbeing, stability, and risk, potentially making those accepted into such services at higher risk of community failure from the outset. The current findings point to specific covariates which are associated with lower survival rates in the community, highlighting those sub-groups of the OPDP population most in need of support. To increase the length of time that released offenders safely remain in the community, it is suggested that these groups be the focus of future intervention services.

Both the current study and previous literature suggest that multimodal approaches, including support for housing, substance misuse, employment, and risk, are most effective in prolonging release. It is suggested that risk and personality complexity indicators, in particular, are used

to identify the cohort of men who are most likely to require multimodal support following release. Whilst prolonging release is an admirable aim, it is recognised that interventions can also prevent or reduce the severity of reoffending through recall and reconviction decisions.

The challenge for services is to determine the type of support required for men with significant histories of childhood trauma, as this is a clear predictor of shorter survival times in the community, as measured by the childhood difficulties covariate and suggested by the self-harm / suicide attempts covariate (Bruffaerts et al., 2010). The current IIRMS specification (HMPPS & NHS, 2021) describes the need to adopt an individualised approach to supporting complex offenders on released from custody, which encompasses the main aspects of trauma informed services: trust, safety, choice, collaboration, and empowerment (Knight, 2019). The question remains whether services can consistently adopt a trauma informed approach with such complex men or whether more attention needs to be paid to delivering trauma informed care in a criminal justice context. It should also be acknowledged that the length of time required to successfully support men with trauma histories is likely to be significant and that return to custody events are likely to continue in the interim. Thus, examining the longer-term impact of community intervention services, with potentially multiple return to custody events, is necessary to fully understand the real-world outcome for this group of complex men.

Despite the helpful findings of the current study and the important discussion points raised, significant areas for further research are identified. Smaller scale research, including RCTs, where it is possible to limit the interventions offered to control groups will strengthen the methodology of future research. Although offence type (violent / other) was neither a univariate nor multivariate predictor of outcome, it may be helpful for future research to focus on specific categories of offence type to establish the nature of support that may be required. It would be helpful to examine the longer-term impact of community intervention services, as the current study uses a 12-month follow up period and is limited to one return to custody

event. Future research should also focus its attention on harm reduction. Whilst high-quality data on harm reduction was not available in the current study, future research should focus on both the frequency and severity of further offending, to explore the notion that interventions may be successful in reducing the level of harm caused by previously violent, high-risk, complex offenders. Attention should also be paid to the reasons that cases are recalled, particularly given the finding that approximately 80% of parole violations are related to technical, rather than criminal, acts (Antonio & Crossett, 2017). The finding that being white had a significant univariate and multivariate negative effect on survival in the community for the IIRMS population requires further exploration. Explanations for these results are not possible within the current study and further research should address this issue. A detailed examination of whether ethnicity plays a role in outcomes within IIRMS is required, as the current findings look only at the combined IIRMS and control population. Qualitative research is also required to examine the experiences of white and non-white service users and their journeys to recidivism and desistance. This may allow interventions to deliver more culturally sensitive services that increase the likelihood of community survival for all ethnic groups. Furthermore, qualitative research building on themes from Ryan et al. (2019) and Craissati et al. (2021) are essential in identifying the nature of support that has the most significant impact on wellbeing and risk.

In combination, the literature review and research have raised concerns about how progress is accurately measured, alongside the blunter and binary recidivism measures used in previous research. The following chapter provides a detailed analysis of the CORE-OM, a user-friendly measure of global distress and risk, in order to contribute to this debate. Despite not being used as an outcome measure in the empirical study (see Chapter 1 for further discussion), the CORE-OM has the potential to add nuance to the outcome measures traditionally used in forensic research. By understanding the progress, or otherwise, that service users make during

interventions in terms of wellbeing and emotional distress, it is argued that services can adapt and develop their interventions to meet the specific needs of clients.

CHAPTER 4

Critique of the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM)

Introduction

The modern field of psychology recognises that individual difference plays a central role in how people adapt to changing circumstances, such as release from custody, often with significant implications for mental and physical health (Segerstrom & Smith, 2019). Psychological measurement aims to examine these individual differences to provide professionals with information from which to make informed clinical decisions (Urbina, 2014). The complexity of the human mind makes the objective measurement of individual difference a significant challenge, thus we accept that this process provides an estimate at best. That said, well designed and validated psychometric assessments are thought to provide a better estimate of difference than subjective observation (Coaley, 2010).

Within mental health and psychological services, outcome measures are required to inform clinicians about the wellbeing of clients, the efficacy of treatment and the potential need for further intervention or support. Over the last two decades, the Department of Health has committed to implement routine outcome measures that are applicable to a wide range of client groups (Fonagy et al., 2004). Whilst it would be impossible to develop an outcome measure that accounts for the whole range of presenting difficulties within mental health services, a number of measures have been developed which focus on a broad spectrum of common concerns. The Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM; Evans et al., 2000) is one of these tools and is the subject of the current critique.

The CORE-OM was designed in response to calls for a core battery of measures, applicable to all users of mental health and psychological services (e.g. Waskow, 1975). At the time of development there were a vast range of measures being used in the UK and US to assess specific areas, such as depression, anxiety, and personality, although many of these measures were only being used by one service, making comparisons impossible (Froyd et al., 1996;

Mellor-Clark et al., 1997). Although previous attempts had been made to develop robust core batteries for the assessment of anxiety disorders, mood disorders and personality disorders (Parloff, 1997), criticisms were levelled in relation to their length, readability, generalisability to other cultures and lack of referential and normative data, among others (Barkham et al., 1998). The CORE-OM was developed to provide a standardised initial assessment and outcome battery which could be used to evaluate the progress, or otherwise, of service users and services, as well as providing valuable data for research purposes. It was designed for use with supplementary tools assessing specific conditions, such as psychosis.

The CORE-OM has been verified in a general population sample (Connell et al., 2007), in large primary care samples (Evans et al., 2003), in secondary care samples (Barkham et al., 2001), in a combined primary and secondary care sample (Barkham et al., 2005), with an older adult population (Barkham et al., 2005) and with service users experiencing Borderline Personality Disorder (Whewell & Bonanno, 2000). It has become a common tool in both counselling and psychological therapy services (Barkham et al., 2006) and is increasingly being used in secure or community forensic settings (Perry et al., 2013). The current critique aims to evaluate the CORE-OM's scientific properties through an assessment of the tool's validity, reliability, and alignment with routinely used supplementary assessments. It will also focus on the utility of the CORE-OM with forensic clients, drawing on literature evaluating risk in the context of attachments (Bowlby, 1977), the Good Lives Model (Andrews et al., 2011) and Theory of Needs (Maslow, 1943).

Selection of the CORE-OM for Examination

The CORE-OM was selected for examination as it provides a potential alternative, or ideally a companion, to existing measures of success with forensic clients. Traditionally, services have

relied upon binary measures of recidivism to determine whether a service or individual has succeeded on release from prison. The major issues with this approach are that services and service users are not credited for progress made in terms of time spent in the community, increases in wellbeing, reductions in emotional distress or reduced harm when comparing past with subsequent offences. Whilst some of the research discussed in Chapter 2 used different measures of progress, all studies, including the original research in Chapter 3, focused heavily on binary measures of recidivism. This leaves a significant gap in our understanding of how we measure the nuances that exist in progress through interventions, and it is believed that the CORE-OM is a measure that can potentially fill this knowledge gap. It is also anticipated that routine collection of wellbeing and global distress information will enable services to establish which questions correlate with both positive and negative outcomes, thus allowing interventions to adapt to meet individual needs.

Overview of the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM)

The CORE System was designed as an evaluation, audit and outcome benchmarking system for psychological therapy services (Core System Group, 1998). Whist acknowledging the practitioner completed evaluation and audit sections, this critique focuses on the third component of the system, the CORE–Outcome Measure (CORE-OM). This is a short patient-completed measure of global distress that can be administered multiple times before, during and after therapy and is validated for use on its own (Core System Group, 1998).

The CORE-OM is a 34-item, self-completion questionnaire comprising direct questions using simple language. Each item is scored on a five-point scale ranging from 0 (not at all) to 4 (most or all the time). Responders are asked to rate their answer to each question based on how they

have been over the last week. Half of the items refer to low-intensity problems (e.g., I have difficulty getting to sleep or staying asleep) and half to high-intensity problems (e.g., Unwanted images or memories have been distressing me). Eight items are described positively (e.g., I have felt warmth and affection for someone) and are reverse scored, with the remainder referring directly to problems. Low- and high-intensity problems were included to increase the scoring range and therefore the tool's sensitivity to change (Evans et al., 2002).

The CORE-OM is arranged across four domains; *subjective well-being* includes questions regarding feelings about self and optimism for the future (four items); *problems/symptoms* incorporate questions related to anxiety, depression, physical problems and trauma (12 items), *life functioning* comprises questions on general day to day functioning, close relationships and social relationships (12 items) and *risk* focuses on both risk to self and risk to other (six items). Overall and domain scores can be pro-rated for up to three missing items in the overall assessment and up to one missing item in each domain. Once an overall average score is calculated (total score divided by number of completed items, assuming fewer than three missing items), a Clinical Score is generated by multiplying by 10, to ensure that clinically meaningful differences are expressed as whole numbers (Connell et al., 2007). This results in a range of scores between 0 and 40 (Leach et al., 2005). The assessment comes with a comprehensive user manual providing both technical and administration information for the CORE System, including details of the informed consent process. Table L1 in Appendix L contains the dimension breakdown of the CORE-OM.

Characteristics of the CORE-OM

Self-Report

The CORE-OM is a self-report measure which is used to complement the two practitioner-completed measures within the CORE System. However, as the CORE-OM is routinely used alone, it is important to note the strengths and weaknesses of self-reported information. Often the most accurate way of gauging the experiences of a client is to ask them directly (McDonald, 2008). Whilst clinical interviews adopt this approach, they can often induce anxiety or be viewed as threatening or confrontational by clients, particularly by forensic clients and when questions are personal (Ramsden, 2018). Self-report measures avoid some of the relational issues that clinical interviews suffer from, thus should be a relatively accurate method of understanding a client's recent experiences. However, it should also be noted that clients may provide distorted responses, such as under-playing the severity or recency of symptoms or behaviours to create a positive impression from administrators. This could be to avoid detention or treatment or to reduce distressing emotions (Williams et al., 2019). Conversely, clients may over-play the recency or severity of symptoms to elicit further care or to avoid discharge from services (Giromini et al., 2022).

Whilst it is acknowledged that all assessments come with measurement error and should only represent an estimate of the domain being measured, the issues described above regarding conscious or unconscious bias in respondents should be carefully considered when interpreting scores from the CORE-OM. It is recognised that there are statistical procedures that can be used to correct for socially desirable responding, such as those described by Saunders (1991). However, it should also be noted that, used as a pre- and post-intervention assessment of progress, of primary interest is the difference between CORE-OM scores, as opposed to

absolute values. This is likely to reduce, although not eliminate, the impact of socially desirable responding.

Psychometric Properties of the CORE-OM

For a psychometric test to be considered 'good' and therefore usable in an applied setting, it requires several key features. It needs to use an interval scale, meaning that the test needs to place individuals in a particular order depending on their score and that this score also indicates the difference between individuals. The test needs to be reliable, meaning that it needs to produce approximately the same result for the same person each time, accounting for the measurement error. It needs to be valid, meaning that it accurately measures what it claims to measure. It needs to demonstrate consistency with other tests known to be assessing the same construct and it needs to accurately differentiate between clinical and non-clinical samples. Finally, a good test needs to have appropriate norms in order that scores from individuals can be understood in the context of the range of scores we would expect from the population being studied. These areas are examined in detail with reference to the CORE-OM below.

Scale

As is described above, completion of the CORE-OM results in a Clinical Score ranging from 0 to 40 (Leach et al., 2005). Although this includes a score of zero, it cannot be said to be a 'true' value of zero, as we are unable to be confident that such a score represents a *complete* absence of the item being measured (Coaley, 2010). As we cannot be confident what a 'true' zero is in relation to these constructs, the CORE-OM is interpreted as an interval scale. This means that individual scores can be said to be *higher* or *lower* than others and assumes that equal numerical differences represent equal changes in the construct being measured. This

also allows the same individual to be measured for improvements or declines at different times in the therapeutic process.

Acceptability

As a self-report measure, of fundamental concern to services using the CORE-OM is the acceptability of the tool to clients. An approximation of acceptability is both the response rate and the pattern of omitted items. Evans et al. (2002) report a complete data response rate of 91% from their non-clinical sample and 80% from their clinical sample. While this difference was statistically significant ($p<.001$), when pro-rating missing data where three or fewer items were missing, the populations were comparable, with 98% of the non-clinical sample returning data and 97% of the clinical sample. The overall omission rate was 1.7%. Connell et al. (2007), Byrne et al. (2010) and Barkham, Gilbert et al. (2005) all report good acceptability data with both non-clinical and clinical samples. Barkham, Culverwell et al. (2005) established significantly lower completion rates with older adults compared with the normative data used by Evans et al. (2002). This either suggests that the tool may be less acceptable with this group when compared with the normative sample or that older adults experience similar difficulties across a range of self-completion tools (e.g., Hayes et al. 1995).

Reliability

Reliability refers to the consistency and repeatability of a test. There follows an examination of the internal reliability, or internal consistency, and the external reliability, or test-retest reliability, of the CORE-OM.

Internal Consistency

Cronbach's alpha (Cronbach, 1951) is the standard method of measuring internal consistency by measuring the proportion of variance that is covariant between the items on a test. Cronbach's alpha scores range between 0-1. A low alpha value suggests that items are poorly correlated to each other thus not tapping a psychological construct related to individual difference. A very high alpha value, approaching one, suggests that there are too many items which are semantically equivalent, thus not adding to the discriminatory value of the test (Byrne et al., 2010). Examining each domain separately, Evans et al. (2002) report alpha coefficients between .75 and .94 for each individual domain, the three combined non-risk related domains and all items within the CORE-OM (see Table 10, below). Confidence intervals were small enough to suggest precise estimates due to the large sample size.

Table 10

Coefficient Alpha (95% CI) Denoting Internal Consistency For Non-Clinical and Clinical Samples (Reprinted from Evans et al., 2002)

Domain	Non-clinical (n=1009)	Clinical (n=713)
Subjective well-being (4 items)	.77 (.75-.79)	.75 (.72-.78)
Problems/symptoms (12 items)	.90 (.89-.91)*	.88 (.87-.89)*
Functioning (12 items)	.86 (.85-.87)	.87 (.86-.88)
Risk (6 items)	.79 (.77-.81)	.79 (.77-.81)
Non-risk items (28 items)	.94 (.93-.95)	.94 (.93-.95)
All items (34 items)	.94 (.93-.95)	.94 (.93-.95)

**p<0.05 (significantly higher α in non-clinical sample)*

No significant differences were found between alpha coefficients in the clinical and non-clinical samples, except for the problem/symptom domain where the non-clinical sample had a significantly higher coefficient (both remained within acceptable limits). Connell et al.

(2007), Barkham et al. (2005) and Byrne et al. (2010) support these findings by reporting alpha coefficients between .70 and .89 for each domain, although higher scores in the problem/symptom and functioning domains suggest that there may be some redundant items within these domains which add little to the sensitivity of the test.

Test-Retest Reliability

Test-retest reliability measures the repeatability of a test across time and is assessed by testing the same person, or people, at two different times and calculating the correlation between scores. Coefficients range between 0-1, with 0 indicating no reliability and 1 indicating perfect reliability. Coefficients above .75 have traditionally been considered to demonstrate excellent reliability (Fleiss, 2011). Reliability depends on the theoretical consistency of the trait or state being measured and is therefore subject to variability with the CORE-OM, when assessing the changeable concept of psychological distress. To mitigate some of this natural variability, it is recommended that the time lag between test one and test two is kept short, ideally between two and 14 days, in order that significant changes in psychological state are minimised (Streiner et al., 2015). Evans et al. (2002) tested a small non-clinical sample of students one week apart and obtained test-retest reliability coefficients for each domain of between .64 and .91 (see Table 11). Included in Table 11 are test-retest reliability data from one of the many translations of the CORE-OM into other languages, in this case Swedish (Elfstrom et al., 2013). Tested between seven and 18 days apart, Elfstrom et al. (2013) established correlation stability coefficients which were marginally lower than Evans et al. (2002) but followed a similar pattern. Using clients in a therapeutic forensic setting, McCloskey (2001) assessed the reliability of the CORE-OM by testing 54 prisoners two weeks apart. Lower correlation stability coefficients were found across all domains, which the author suggests is because test

one was conducted shortly after arriving in a new and very different therapeutic environment and test two was conducted two weeks later when residents had begun to settle in. This may be particularly pertinent for the risk domain with forensic clients, as it contains items that relate to harming others and harming self, which may be expected to fluctuate during times of transition. The lowest correlation stability coefficient across the three studies related to the risk domain (.64, .64 and .58, respectively), which Evans et al. (2002) explain as being due to the small number of items (n=6) and their reactive nature. The highest correlation stability coefficient established by Evans et al. (2002) and Elfstrom et al. (2013) (.91 and .86, respectively) was obtained in the combined domains minus the risk domain, although McCloskey (2001) established the highest correlation stability coefficient using all items in the tool. Whilst not conclusive, these findings indicate that there may something measurably different about the risk domain that requires further exploration.

Table 11

Test-Retest Stability in Non-Clinical Samples (Adapted from Evans et al., 2002, Elfstrom et al., 2013 and McCloskey, 2001)

CORE-OM Domain	Spearman's *p* *		
	Evans et al. (2002) n=43	Elfstrom et al. (2013) n=70	McCloskey (2001) n=54
Well-being	.88	.80	.68
Problems	.87	.80	.66
Functioning	.87	.81	.62
Risk	.64	.64	.58
Non-risk items	.91	.86	.72
All items	.90	.85	.74

**Correlation stability coefficient*

Validity

Put simply, the validity of a test is the extent to which it measures what it purports to measure. Validity, therefore, affects the accuracy of the inferences that are made from test results. There is no one measure of validity which is considered superior and different measurements contribute to an overall assessment of validity (Coaley, 2010). Below is an examination of relevant assessments of the CORE-OM's validity.

Face Validity

An assessment is said to have face validity if it appears to measure what it claims to measure (Anastasi & Urbina, 1997), although as it is not underpinned by established theory (Fink, 2010), face validity is often considered a controversial measure (Sartori & Pasini, 2006). The fact that the CORE-OM has undergone rigorous research and has become one of the most widely used assessments for psychological distress within counselling, psychotherapy and, increasingly, in forensic settings, suggests that clinicians have confidence in its face validity. Acceptability data indicates that a high percentage of respondents are willing to complete the CORE-OM and that missing item rates are generally low. As a subjective measure of validity, the CORE-OM appears to have acceptable face validity.

Concurrent Validity

Concurrent validity is a measure of the relatedness of the test in question to other established tests, which are known to measure the same concept. Smith and Smith (2005) suggest that a correlation of over .55 between measures provides good evidence of concurrent validity. Connell et al. (2007) compared the CORE-OM with the Clinical Interview Schedule-Revised

(CIS-R; Lewis et al., 1992) and Leach et al. (2005) compared it with the Health of the Nation Outcome Scale (HoNOS; Wing et al., 1998), with findings indicating a proportionate share of the variances of .77 and .50, respectively. The risk domain of the CORE-OM shared a proportion of .57 of the variance, with the behavioural problems domain of the HoNOS. This is notable, as the risk domain of the CORE-OM is considered to have the lowest test-retest reliability of all domains, whilst also having good concurrent validity with the HoNOS' own risk domain. This suggests that both tools are able to identify common features of people presenting with severe risk related problems.

Evans et al. (2002) summarise a body of research comparing the CORE-OM with other measures of psychological well-being and symptomology using clinical samples. As Table 12 demonstrates, the CORE-OM correlates well with a range of different assessments including the Beck Depression Inventory-I and -II (BDI; Beck et al., 1961, 1996), Beck Anxiety Inventory (BAI; Beck et al., 1988), Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1983) and the General Health Questionnaire (GHQ; Goldberg & Hillier, 1979). Correlations were generally highest with conceptually similar measures such as the SCL-90-R (.88), the Brief Symptom Inventory (.81; BSI; Derogatis & Melisaratos, 1983) and the BDI-I and -II (.85 and .81, respectively). Although confidence intervals for the BDI-II suggested that estimates of correlation were imprecise due to a small sample size (n=29), a later study by Cahill et al. (2006) indicated excellent correlation between the CORE-OM and BDI-II with a larger sample (n=77). It is noteworthy that the risk domain of the CORE-OM showed the lowest correlation of all domains with many of the associated measures, suggesting that risk may not be an area effectively captured by other general psychological distress measures. However, the risk domain appears highly correlated with the SCL-90-R and the GHQ-D (severe depression) measures and has also been shown by Evans et al. (2002) to be well correlated with clinician's ratings of 'significant risk' in a population of 40 students from a university counselling service.

These findings, supported by the development of conversion tables between the CORE-OM and the BDI (Leach et al., 2006), provide evidence of good concurrent validity with other measures of psychological distress and disturbance.

Table 12

Correlations with Referential Measures in Clinical Samples *(Reprinted from Evans et al., 2002)*

Measure	N	CORE-OM Domain					
		Well-being	Problems	Functioning	Risk	Non-risk items	All items
BDI-I	251	.77	.78	.78	.59	.84	.85
BDI-II	29	.79	.74	.78	.32	.83	.81
BAI	218	.56	.68	.55	.39	.65	.65
BSI	97	.63	.76	.71	.62	.79	.81
SCL-90-R	34	.68	.87	.79	.83	.85	.88
GHQ-A	69	.43	.60	.44	.30	.56	.55
GHQ-B	69	.55	.61	.57	.30	.64	.64
GHQ	69	.67	.66	.65	.56	.72	.75
GHQ-C	69	.60	.52	.60	.44	.62	.63
IIP-32	246	.48	.58	.65	.45	.64	.65
GHQ-D	69	.63	.47	.55	.69	.58	.63

BAI, Beck Anxiety Inventory (Beck et al, 1988); BDI, Beck Depression Inventory (Beck et al, 1961, 1996); SCL-90-R, Symptom Checklist-90-Revised (Derogatis, 1983); BSI, Brief Symptom Inventory (Derogatis & Melisaratos, 1983); IPP-32, Inventory of Interpersonal Problems – 32-item version (Barkham et al, 1996); GHQ, General Health Questionnaire, 28-item version (Goldberg & Hillier, 1979); A, somatic symptoms; B, anxiety and insomnia; C, social dysfunction; D, severe depression.

Determining Differences Between Clinical and Non-Clinical Samples

One of the main requirements of any assessment is that it is successfully able to discriminate between the clinical population it is designed to be used with and non-clinical populations. As is shown in Table 13, The Core System Group (1998) report significant differences between clinical and non-clinical populations across all four domains of the CORE-OM, as well as for the three combined non-risk domains and for all combined domains. Confidence intervals are small, indicating precise estimates, with differences of over one point on a 0-4 interval scale

for all domains except risk, which had a difference of less than one point. This suggests that the CORE-OM is able to successfully distinguish between clinical and non-clinical populations, making it a useful decision-making tool with regard to the targeted offering of treatment or support.

Table 13

Means and Standard Deviations for Clinical and Non-Clinical Samples (Reprinted from Core System Information Management Handbook; Core System Group, 1998)

Dimension	Non-clinical (N=1084)		Clinical (N=863)		Difference	
	Mean	S.D.	Mean	S.D.	95% C.I.	*p* *
Well-being	0.91	0.83	2.37	0.96	1.38 to 1.53	<.001
Problems	0.90	0.72	2.31	0.88	1.33 to 1.48	<.001
Functioning	0.85	0.65	1.86	0.84	0.95 to 1.09	<.001
Risk	0.20	0.45	0.63	0.75	0.38 to 0.49	<.001
All non-risk items	0.88	0.66	2.12	0.81	1.18 to 1.31	<.001
All items	0.76	0.59	1.86	0.75	1.04 to 1.16	<.001

**p values for Mann-Whitney test*

Factor Structure

The CORE-OM is organised across four domains, or factors: well-being, symptoms/problems, functioning and risk (Barkham et al., 2001). Factor analysis enables an exploration of whether each factor taps a unique concept or whether multiple factors identify the same or similar concepts. High correlations were found between domain scores by Evans et al. (2002), using both clinical and non-clinical populations, although the proportion of variance shared was lower for the risk domain (between .33 and .60). The proportion of variance shared across the remaining three domains was high (between .73 and .78), which was anticipated given the close associations between well-being, symptoms/problems and functioning. This suggests that within the original four factor structure, three of the factors may tap a similar concept. Lyne et al. (2006) suggest that the CORE-OM should be arranged across two factors: psychological

distress (combining the well-being, symptoms/problems and functioning domains) and risk. When examining the psychometric properties of the CORE-OM in a forensic therapeutic setting, Byrne et al. (2010) found similar results, with high levels of homogeneity across the three non-risk domains. Furthermore, there was little support for the four-factor model proposed by Barkham et al. (2001), with only the risk items loading strongly onto one of the factors during confirmatory factor analysis.

These studies suggest that the CORE-OM can measure general psychological distress and risk but is perhaps less capable of differentiating subtleties within the former category. Having said this, Byrne et al. (2010) conclude that the CORE-OM is an acceptable and reliable measure to use in a forensic therapeutic setting, albeit with some ongoing concerns around the factor structure of the tool.

Population Validity

Normative data on which interpretation of scores is based is reported in the Core Systems User Manual (1998), as well as in Evans et al. (2002). The non-clinical data came from three sources; the first of which included 691 students studying different courses at a British university; the second included 55 students from a different British university taking part in a test-retest study; and the third was a sample of convenience including non-clinical staff, friends, and family of the CORE-OM research team. In total, the non-clinical sample included 1,106 people. The clinical data came from 21 different sites, the majority of which were within the NHS, but also included one student counselling service and one staff support service. Only data from pre-treatment or first treatment sessions were used to avoid any effects of treatment. In total, the clinical sample included 890 people.

Whilst this provides an adequate sample size on which to base normative data, the large number of sites providing data means that there may have been wide variations in how the tool was administered and the conditions under which it was completed.

Reliable and Clinically Significant Change

An important characteristic of a test is that it is reliably able to determine clinically significant change in individuals due to treatment. Evans et al. (2002) set out to determine whether the CORE-OM was sensitive enough to measure reliable and clinically significant change using three clinical samples (primary care counselling service, student counselling service and remaining NHS psychotherapy/counselling services). Based on Jacobson and Truax's (1991) work, reliable change is that which only 5% of the population show if change is down to unreliability of measurement, whereas clinically significant change is the movement of a score from one which is characteristic of a clinical population to one which is characteristic of a non-clinical population. Using their clinical samples, Evans et al. (2002) determined 'cut-scores' from which they could measure the level of reliable and clinically significant change using the CORE-OM. Most patients showed reliable improvement in the primary care counselling service (74%) and the student counselling service (71%), with 49% showing reliable improvement in the NHS psychotherapy/counselling service. Clinically significant improvement was established in 50%, 49% and 41% of patients in each sample, respectively, demonstrating that these patients moved from a score typical of a clinical sample to one typical of a non-clinical sample. In interpreting these results, it is worth noting that patients whose scores resembled the non-clinical sample to start with would not have been able to achieve clinically significant improvement and patients who scored very highly in the clinical range may have demonstrated reliable change but not to the degree of moving into the non-clinical

range. Evans et al. (2002) concluded that these results are supportive of the CORE-OM's ability to sensitively measure reliable and clinically significant change.

Using a forensic population of adult victims of childhood abuse, Byrne et al. (2010) established that 68% of patients demonstrated reliable improvement and 42% showed a clinically significant improvement. Although the sample size was small (n=19), conclusions were drawn that the CORE-OM is a suitable tool to be used with this population and that the service being provided was as effective as counselling and psychotherapy services in achieving reliable and clinically significant change.

Application of the CORE-OM in Forensic Settings

Of particular interest to the current critique is the CORE-OM's increasing use in secure forensic settings to measure treatment outcomes or predict engagement (e.g., Perry et al., 2013; Kotterbova & Lad, 2022). Whilst risk has traditionally been the focus of assessments in secure forensic settings (Fitzpatrick et al., 2010), a range of other considerations have come to the fore over recent years, as a more holistic understanding of risk has developed. This has included our understanding of Attachment Theory (Bowlby, 1977), desistence and Good Lives (Andrews et al., 2011) and the Theory of Needs (Maslow, 1943), all previously described in Chapter 3. The CORE-OM taps many of these areas within the subjective well-being, symptoms/problems, functioning and risk domains, indicating that it may be a useful tool to use with forensic populations.

The risk sub-scale of the CORE-OM has been found to correlate highly with experienced clinician's views of patients with Borderline Personality Disorder's risk to self and other (Whewell & Bonanno, 2000). Whilst it is accepted that the sample in the study by Whewell and Bonanno was an out-patient non-forensic population, it is also known that a significant

minority of prison and secure mental health residents experience traits of BPD (Lovell & Hardy, 2014), suggesting that these findings may also apply to a forensic population. McCloskey (2001) validated the CORE-OM on a population of male offenders in a secure therapeutic forensic setting, finding acceptable internal consistency (alpha between .62-.95), test-retest reliability (see Table 11) and preferential acceptability to Evans et al. (2002). Perry et al. (2013) established that the CORE-OM is acceptable to both patients and staff across high, medium, and low secure mental health services. Using outcomes from questionnaires, they also found that the CORE-OM was feasible and potentially useful to administer with a forensic population. However, there are several caveats to consider, including that response rates were low (25%), residents with a primary diagnosis of personality disorder or learning disability were excluded from the study and the sample was exclusively male and predominantly identified as 'White British' (88%). Whilst the response rate and male only sample is not unusual for research in forensic populations, the fact that the population excluded some significant minority groups means that results cannot be generalised to the wider forensic population. Further research including some of these groups will be helpful and is increasingly possible, particularly as IIRMS have started to use the CORE-10 (a ten-item version of the CORE-OM) as part of their pre- and post-intervention battery (HMPPS & NHS, 2021). Byrne et al's (2010) validation of the use of the CORE-OM with a population of adult victims of childhood abuse is helpful in broadening the application of the tool but focuses on victims of crime, as opposed to perpetrators. Whilst it is not possible to generalise these findings to a forensic population, we know that a high proportion of offenders' experience significant childhood trauma (e.g., DeLisi et al., 2019; Fox et al., 2015), suggesting that it would be helpful to develop further evidence for the use of the CORE-OM with a forensic population.

The studies described above provide evidence for the appropriate use of the CORE-OM in forensic settings, although further research is required to support its use with specific sub-

populations. Whilst McCloskey's (2001) validation of the CORE-OM will have included a significant proportion of men with personality disorder (Taylor & Trout, 2013), this population had voluntarily agreed to engage in treatment, potentially making the acceptability of the tool higher than in other forensic PD populations. Several other studies used non-PD populations from secure mental health services (Perry et al., 2013), female only participants (Kotterbova & Lad, 2022; Whewell & Bonanno, 2000), community sentenced offenders (Kotterbova & Lad, 2022) or non-forensic populations (Whewell & Bonanno, 2000), meaning that the use of the CORE-OM with a mainstream male population of custodial sentenced offenders with personality disorder traits has not been fully explored. Despite this, the tool is widely used in prisons and secure mental health settings (e.g., McIntosh et al., 2018), including with clients with personality disorder (e.g., Macinnes et al., 2016). Considering the difficulties with engagement that personality disordered forensic clients experience (Shuker, 2010), it may be expected that the acceptability of the CORE-OM would be lower than with other populations. Given the presence of the risk domain, which is particularly pertinent to forensic populations, it is important that the CORE-OM is specifically validated with this group. Furthermore, data regarding reliable and clinically significant change is required with personality disordered forensic clients, as we know that therapeutic change with this group can be slow (Bateman et al., 2015).

Conclusions

The CORE-OM provides a general measure of psychological distress and was born out of a desire to develop a core battery of assessments for use in mental health and psychological services (Barkham et al., 1998). Studies have demonstrated the CORE-OM's acceptability, with good response rates and low item omission for clinical and non-clinical populations

(Evans et al., 2002), as well as in a forensic therapeutic setting (McCloskey, 2001). However, it should be noted that in a forensic mental health setting, Perry et al. (2013) established a response rate of only 25%, raising some concerns about the CORE-OM's acceptability in this environment. Reliability of the CORE-OM has been assessed, with acceptable internal consistency across both clinical and non-clinical populations and test-retest reliability with a non-clinical population (Evans et al., 2002). Test-retest reliability data comes from a small sample (n=43) of students, and it would be helpful to repeat this process with a larger sample including both clinical and non-clinical participants. These data suggest that the CORE-OM is both an acceptable and reliable tool, although caution should be exercised when using it in more specialist settings.

The CORE-OM has evolved into one of the most widely used measures of psychological distress in counselling and psychotherapy settings, suggesting good face validity. Concurrent validity has been established with correlations greater than 0.55 between the CORE-OM and BDI, BAI, SCL-90-R and GHQ (Evans et al., 2002). The CORE-OM is also able to discriminate between clinical and non-clinical samples across all four domains. This evidence indicates that the CORE-OM meets Kline's (1986) characteristics of a good test.

The research base supporting the use of the CORE-OM is relatively strong, involving large clinical and non-clinical samples. Whilst these samples are likely to have included people experiencing Autistic Spectrum Conditions, Personality Disorders and Learning Disabilities, no specific validation has been conducted with these groups, suggesting that this would be a valuable area for future research. Whilst the CORE-OM is routinely used across many forensic settings, only a small handful of studies have validated the CORE-OM's use with this population. Further support for the use of the CORE-OM with perpetrators of crime is needed and it would be helpful to determine the concurrent validity of the risk domain against forensic risk assessments such as the HCR-20v3 (Douglas et al., 2013) and the RSVP (Hart et al., 2003).

Within the forensic population, it would also be valuable to establish whether the CORE-OM is an acceptable, reliable and valid tool to be used with those diagnosed with a Personality Disorder.

Whilst the CORE-OM meets the requirements to be considered a strong psychological test, the factor structure raises some concerns. Studies suggest that the test may not be sensitive enough to discriminate between the first three factors accurately. Instead, it is suggested that the CORE-OM may more accurately provide a general measure of distress (combined well-being, symptoms/problems and functioning domains) and a general measure of risk (to both self and other). It may be helpful for future research to consider a simplification of the factor structure of the CORE-OM. Further research is also required to determine whether the six-item risk domain could be used in isolation to provide a quick approximation of current risk, particularly with forensic populations.

The CORE-OM has a strong evidence base which confirms it as an acceptable, reliable, and valid tool to assess general psychological distress in both clinical and non-clinical populations. Perhaps the strength of the tool as a general measure of distress and its use with a range of clients has, over time, also become its weakness. The CORE-OM's increasing use in specialist settings renders it vulnerable to criticism due to the lack of specific validation with these populations. Future research may help prevent its use in inappropriate settings and will increase our confidence when using the tool with specialist groups, such as offenders.

Findings from chapters 2 and 3 highlight the emphasis placed on recidivism measures to determine the efficacy of community interventions with this complex group of release offenders. The current chapter demonstrates that the use of a global distress measure is not only appropriate with this group but also adds nuance to our understanding of an individual's progress through treatment. The following chapter aims to bring together key findings from

the preceding chapters to create a coherent understanding of who is most in need of such services, which interventions appear to be most efficacious and how progress through services can be best measured. Suggested directions for further research are made to build on our understanding of these three critical areas.

CHAPTER 5

General Discussion

Aims of the Thesis

The purpose of this thesis is to examine community services available for high-risk, complex, violent, male offenders on release from custody. Specifically, this thesis examines the international literature focused on existing services for this group, to identify approaches that reduce recidivism and prolong release in the community. Existing UK intervention and risk management services for this group were then explored to understand their efficacy and whether sub-groups of this population benefitted from such services. As it was established that progress through interventions was typically measured by unsophisticated recidivism data, attention was then paid to the CORE-OM, a global measure of distress. A critique of the use of the CORE-OM in general, and specifically with forensic populations, was presented with a view to better understanding the range of measures required to accurately assess the efficacy of community intervention services.

The systematic literature review in Chapter 2 examines the international literature on community interventions for high-risk, complex, violent, male offenders on release from custody. This has allowed for a comprehensive understanding of interventions that have been successful in both reducing recidivism and prolonging release in the community. The empirical study presented in Chapter 3 compliments these findings by examining current UK interventions for this group, allowing for a greater understanding of efficacy. Chapters 2 and 3 also offer significant insights into the efficacy of interventions for sub-populations of this group. This increases our understanding of the effects of age, sentence type, risk level and personality disorder traits on the chances of survival in the community. Finally, the examination of the CORE-OM in Chapter 4, alongside the original research in Chapter 3, builds on our understanding of how to effectively measure the success, or otherwise, of intervention services in the community. In summary, the aims of the thesis were met and further areas for research and discussion are identified.

The remainder of the discussion is organised around three critical areas: who is most in need of high-intensity community intervention services, what type of intervention or support services are most effective, and how is the efficacy of such services best measured. Conclusions are then drawn about how these findings can be incorporated into the current IIRMS specification.

For Whom Are Community Intervention Services Most Needed?

This section brings together the findings from this thesis which identify sub-groups from the high-risk, complex, violent, male offender population for whom community intervention appears most necessary. This section appears first, as it is the initial question that needs addressing when providing any intervention service in any setting - for whom is the service most needed? The systematic literature review presented in Chapter 2 reports on several services which appear to have been successful in reducing recidivism and prolonging release in the community (Antonio & Crossett, 2017; Clark, 2015; Lattimore & Visher, 2009). Of these, only one (Antonio & Crossett, 2017) was able to report on a sub-group of their population. They established that as age increased, there was a significant reduction in the risk of reincarceration. The research presented in Chapter 3 established similar results, independent of the intervention offered, confirming that as age increases the necessity for community intervention decreases. In combination, these findings suggest that community intervention services for this complex group of released offenders may be most needed by younger men.

The research in Chapter 3 identified both univariate and multivariate predictors of shorter survival times in the community, independent of the intervention offered. Although the three groups produced slightly different results, general findings were similar. Perhaps unsurprisingly, higher risk scores on RoSH, OGRS1 and OGRS2 were significant predictors

of shorter survival times in the community. Determinate sentenced offenders were also found to be returned to custody sooner, suggesting that men with less serious convictions struggled more in the community. Given that most determinate sentences are significantly shorter than indeterminate sentences, it is possible that short stays in custody cause instability without any significant deterrent effect on release. Shorter stays in custody also limit the opportunity to benefit from treatment and also reduce the generally positive effect of maturation. It was also established that being white (vs. non-white) was a significant predictor of shorter survival times in the community. Without further analysis of the data, which was not possible in the current study, an explanation for this finding is not currently offered. However, further research into IIRMS should include a focus on ethnicity to see if this finding can be replicated and, if so, to explore some of the potential reasons underpinning this finding.

Within the combined intervention group, independent of intervention offered, a history of self-harm/suicide attempts, and challenging behaviour were both predictive of shorter survival times in the community. When combining the OPDP screening items into a single complexity score (0-5), it was established within the IIRMS group that scores of 3 or more significantly predicted shorted survival times in the community when compared to men with lower complexity scores.

These findings offer valuable insight into sub-groups of the high-risk, complex, violent male offender population who appear to struggle most when released into the community. Regardless of intervention offered, it is younger white men, with higher risk scores, on determinate sentences who appear to require additional support on release. Furthermore, men with a history of self-harm/suicide attempts and challenging behaviour appear to fair worse on release, as do men with higher levels of personality complexity, as measured by the combined complexity score. This raises the question of how this data can be used by community intervention services to identify those men who may benefit most from additional support on

release from custody. The variables with the strongest predictive relationship to reincarceration were very high RoSH scores and complexity scores of 5, both either over, or close to, three times more likely to return to prison than men with lower scores. This suggests that these two variables, in particular, could be used by services to identify men who would most benefit from additional support or treatment on release.

Which Types of Intervention Appear to Work Best?

This section summarises findings from the systematic literature review in Chapter 2 and the research in Chapter 3 regarding the type of services that appear to be effective in prolonging release for this complex group of offenders. Several of the studies reviewed in Chapter 2 exclusively involved cohorts of men convicted of sexual offences (Hanson et al., 2004; Hedderman & Sugg, 1996; Procter, 1996), finding limited or no positive results. More recently, services have adopted a person-centred approach, focusing on the needs of the individual (e.g., Antonio & Crossett, 2017; Clark, 2015; Nathan et al., 2019) as opposed to the index offence committed. With more promising results coming from these studies, it is suggested that services take an individualised approach to community management, rather than providing separate services for men with specific convictions. This is consistent with the current IIRMS specification, which describes a formulation led approach and makes no mention of different services for specific offender groups (HMPPS & NHS, 2021). This approach is supported by the research in Chapter 3, which found no statistically significant difference in survival rates between men convicted of violent index offences and those convicted of non-violent index offences, suggesting that offence type plays no role in return to custody outcomes.

Although the literature review in Chapter 2 excluded studies that mostly or exclusively delivered services in custody, it included services that began delivery in custody and continued

into the community. Five studies included a transitional element (Bouffard & Bergeron, 2006; Braga et al., 2009; Clark, 2015; Lattimore & Visher, 2009; Nathan et al., 2019), all of which reported statistically significant or promising results of interventions on recidivism. This is supported by Berghuis' (2018) meta-analysis and is incorporated into the current IIRMS specification, which recognises that in-reach work can improve engagement in the community with this complex group (HMPPS & NHS, 2021). Regarding the research presented in Chapter 3, it is not clear whether each service involved a transitional element, making it impossible to determine how this may have affected survival in the community.

Whilst the IIRMS studied in Chapter 3 were free to innovate and tailor interventions to individuals, there are some findings from the literature review in Chapter 2 which offer insight into the types of services that may be most effective with this complex group of offenders. Both Bouffard and Bergeron (2006) and Clark (2015) aimed to increase released offenders access to services in the community, both studies achieving their aims. Bouffard and Bergeron found that the intervention group were significantly more likely to be referred to additional community services, such as those for housing, education and employment support, and Clark established that as the number of services accessed increased, rates of recidivism decreased. Although Lattimore and Visher (2009) established encouraging results regarding the intervention group's access to housing and employment stability, there was no correspondingly positive effect on recidivism measures. The study presented in Chapter 3 examined five intervention services, although it was not possible to determine how much collaboration there was with other community services. Despite these mixed findings, there is emerging evidence that a multi-modal approach to community support services can have a positive effect on recidivism outcomes for complex offenders.

Findings from this thesis suggest that services supporting this complex group of men in the community should focus on the individual needs of the client, as opposed to the offence for

which they have been convicted. It is also important that the transition from custody to the community forms a central part of any intervention offered, with services beginning in custody aiming to reduce anxiety and increase engagement. Finally, partnership working appears to be an important element of several successful intervention services, suggesting the value of multi-modal service provision in the community. These findings support the new IIRMS specification which outlines the importance of a formulation led approach, the value of transitional services and the need for partnership working (HMPPS & NHS, 2021).

How is Efficacy Best Measured?

Throughout this thesis, the issue of accurately measuring the outcome of interventions is raised. Chapter 2 examines studies that use a range of outcome measures, mainly focused on recidivism, including technical breaches of parole, re-arrest, reconviction and reincarceration. The research presented in Chapter 3 focuses on recall and reconviction, also using a combined measure involving any return to custody. Findings from this thesis suggest that technical breaches of parole should be considered a functionally different measure of recidivism than other more objective measures. Antonio and Crossett (2017) reported that of all parole breaches, 80% were for technical, as opposed to criminal, violations. With more men being recalled than reconvicted in the research presented in Chapter 3, it is highly likely that a significant proportion of these cases were recalled for technical, as opposed to criminal, violations. This suggests that technical breaches of parole may be a better proxy for adherence to rules than for recidivism. The point is made in Chapter 2 that the intended recipients of these services experience high levels of personality disorder traits, often making adherence to rules extremely challenging. Thus, studies reporting a positive effect on technical breaches of parole are likely to be measuring engagement and the ability of the client to work collaboratively with

professionals, rather than recidivism per se. Whilst this does not discount technical breaches of parole as an outcome measure, it is important that services understand the difference between the measures they are using.

Despite more objective measures of recidivism being favoured in most research, there are several weaknesses of these measures described in Chapter 2 and highlighted in the findings of Chapter 3. Whilst much of the research in this area, including that presented in Chapter 3, attempts to match intervention and control groups, it is not always possible to escape the issue of selection bias at the time of referral into a service. Cases accepted for IIRMS, IRiS or similar interventions are likely to be referred at times when risk is increasing or when wellbeing or stability are decreasing. Despite being matched on largely static factors to a control group counterpart, these current factors are likely to make the intervention population more likely to recidivate regardless of any intervention they receive. To avoid the issue of selection bias, a different design of research is required, which is discussed below.

Other factors which affect the accuracy of research reliant on recidivism measures include the possibilities that intervention cases are more closely monitored by services thus more likely to be recalled or, conversely, that they may be provided more leeway when making breach decisions. Outcomes using reconviction data are also likely to represent an under-reporting of events, as there are numerous cases in which an offence is known to have happened but only a recall decision is made, due to the expense and complexity of seeking a conviction. Also raised is the issue of how long change can take with this complex group, the majority of whom experience entrenched traits of personality disorder. People engaged in an intervention may improve on various measures of progress yet may still be returned to custody for a less serious offence or for a technical breach.

This raises the issue of harm reduction and how this can be accurately assessed through community intervention programmes. The examination of the international literature in Chapter 2 indicates how rare it is that research focuses on the reasons people are returned to custody. Whilst Chapter 3 was unable to explore this area in detail, due to insufficient high-quality data, Chapter 4 indicates that the CORE-OM could be a powerful tool to assess progress through community intervention services. This would not only develop our understanding of the effects of interventions on service users' wellbeing and emotional distress but would also allow for longer-term re-evaluations of what to target in treatment. For example, if the functioning domain of the CORE-OM was found to be strongly associated with shorter survival times, interventions could be adapted to focus on living and relational skills, as well as problem solving strategies. In such circumstances, the CORE-OM could also be used as a means of selecting those most in need of community intervention services. What is clear from this thesis is that more nuanced measures of progress are required if we are to fully understand the efficacy of and improve the delivery of community interventions for complex offenders.

Fortunately, the decision to include global distress, wellbeing and hope measures in the assessment of IIRMS has already been taken (HMPPS & NHS, 2021) and this thesis fully supports the range of measures currently proposed.

Suggestions for Further Research

Throughout this thesis, areas for further research are suggested to build a more comprehensive and nuanced understanding of what is effective with this complex group of released offenders. Future research into community intervention services should pay close attention to ethnicity as a potential risk factor for returning to custody. Chapter 3 established that white men were significantly more likely that their non-white counterparts to return to custody sooner,

highlighting the need for this finding to either be replicated or challenged in future research. It is suggested that research focuses on wider groups than simply white and non-white men, in order to better understand differences in recidivism across a range of ethnicities.

Selection bias is a significant issue at the time cases are referred into community services, as accepted cases are often closer to recidivism or recall at this point. Further randomised control trials, such as that conducted by Nathan et al. (2019), should focus on cases accepted for community intervention services and compare those that were randomised into treatment and control groups. Whilst this is the only effective way of eliminating this form of selection bias, it is also recognised that this comes with some ethical concerns when managing high-risk, violent offenders.

Outcome measures are explored above, and it is essential that future research distinguishes between a return to custody for a technical breach and a serious further offence. Although the research presented in Chapter 3 was unable to distinguish between reconvictions for different offences, future research should aim to understand more about the type of subsequent offence committed. This will prevent a dichotomous conclusion around *success* or *failure* from being drawn and will allow for an examination of whether services have a harm reduction effect, rather than simply a recidivism reduction effect.

Future research should compare IIRMS cohorts whose service starts with custodial in-reach with cohorts whose service begins in the community. This thesis suggests that interventions with a custodial in-reach element fair better than those without, and further research will allow this to be tested specifically with an IIRMS cohort. It is also important that services record the specific type of intervention they offer to each person, as the wide variety of interventions, whilst encouraging from a responsivity perspective, makes it difficult to examine which modality of intervention benefits which group of offenders.

Further research is also required to understand whether multi-modal services and partnership working add value to existing interventions. Specifically, it would be helpful to know whether housing, employment, mental health, or substance misuse support increases the efficacy of more traditionally therapeutic services. In line with steps already taken by IIRMS, future research should focus on service user experience when assessing the effectiveness of interventions. Building on the work of Ryan et al. (2019) and Craissati et al. (2021), it is essential that a better understanding of what works for this group is developed, so that the amount of in-reach work, the type of multi-modal support offered, and the specific type of therapeutic intervention delivered can be flexibly applied.

Practical Application of Findings

Several of the findings from this thesis have already been implemented by IIRMS thus lending support to the new IIRMS specification (HMPPS & NHS, 2021). This section summarises these findings and emphasises their importance in the delivery of current services. In addition, new findings are also summarised with suggestions for how these can be implemented in the delivery of current services.

Perhaps the most useful finding from the current thesis is knowing who struggles most in the community following release from custody. The research in Chapter 3 indicates that younger white men, with higher risk scores, serving determinate sentences survive in the community for significantly shorter periods of time than other released offenders. Whilst IIRMS already considers this group as a priority for support, the research also suggests that men with a history of self-harm/suicide attempts and challenging behaviour, as well as overall complexity scores of 3 and above fair worse on release than other groups. Specifically, men assessed as very high RoSH and with a complexity score of 5 have the strongest predictive relationship with

reincarceration. It is suggested that IIRMS should prioritise this small group of released offenders with the aim of reducing the behaviours and concerns that lead to recall or reconviction decisions.

This thesis provides evidence of the efficacy of transitional services that begin their support of offenders in custody and continue this into the community. Whilst the recent IIRMS specification highlights the need for an in-reach element to services, the current findings indicate that this is not just a desirable but an essential element of successful interventions. It is suggested that in-reach support becomes a mandatory aspect of IIRMS, and that consideration is given to developing specific transitional services, such as that offered by the Wales Offender Personality Disorder Pathway (O'Meara et al., 2019). This thesis also highlights the value of multi-modal services and partnership working, which aligns with the IIRMS specifications' emphasis on meeting peoples' core needs (e.g., Deci & Ryan, 2000; Maslow, 1943).

This thesis highlights the importance of the accurate evaluation of community-based interventions. Whilst objective measures of recidivism are helpful, the need for a more nuanced understanding of progress with this complex group of released offenders is recognised. It is acknowledged that progress is not always linear and that returns to custody do not always represent the failure of an intervention. IIRMS have already implemented a range of self-report assessment measures, including the CORE-10, which are supported by the findings of this thesis. However, it is suggested that information on the specific intervention delivered is recorded in order that both recidivism and wellbeing measures can be used in combination to accurately assess the types of interventions that are effective for different service users. It is also recognised that the large-scale collection of wellbeing data can have an effect on intervention design over the longer-term.

References

Abramovaite, J., Bandyopadhyay, S., Bhattacharya, S, & Cowen, N. (2019). Alternatives to custody: Evidence from police force areas in England and Wales. *British Journal of Criminology, 59*(4), 800-822. https://doi.org/10.1093/bjc/azy056

Adekeye, D. S., & Emmanuel, P. (2018). The nexus between early release of inmates and juvenile recidivism. *International Journal of Sociology and Social Policy, 38*(9/10), 837-850. https://doi.org/10.1108/IJSSP-11-2017-0145

Ali, M. S., Groenwold, R. H., Belitser, S. V., Pestman, W. R., Hoes, A. W., Roes, K. C., de Boer, A., & Klungel, O. H. (2015). Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *Journal of Clinical Epidemiology*, *68*(2), 122-131. https://doi.org/10.1016/j.jclinepi.2014.08.011

Ali, M. S., Groenwold, R. H., Pestman, W. R., Belitser, S. V., Roes, K. C., Hoes, A. W., de Boer, A., & Klungel, O. H. (2014). Propensity score balance measures in pharmacoepidemiology: A simulation study. *Pharmacoepidemiology and Drug Safety*, *23*(8), 802-811. https://doi.org/10.1002/pds.3574

American Psychiatric Association. (Ed.). (2022). Diagnostic and Statistical Manual of Mental Disorders (5th ed.- Text Revision). American Psychiatric Publishing. https://www.psychiatry.org/patients-families/personality-disorders/what-are-personality-disorders

Anastasi, A., & Urbina, S. (1997). Psychological Testing. Prentice Hall/Pearson Education.

Andrews, D. A., & Bonta, J. (2003). The level of service inventory revised US norms manual supplement. *Toronto: Multi-Health Systems*.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, *17*(1), 19-52. https://doi.org/10.1177/0093854890017001004

Andrews, D. A., Bonta, J., & Wormith, J. S. (2011). The Risk-Need-Responsivity (RNR) model: Does adding the Good Lives Model contribute to effective crime prevention? *Criminal Justice and Behavior, 38*(7), 735–755. https://doi.org/10.1177/0093854811406356

Antonio, M. E., & Crossett, A. (2017). Evaluating the effectiveness of the national curriculum and training institute's "Cognitive Life Skills" program among parolees supervised by Pennsylvania's board of probation & parole. *American Journal of Criminal Justice*, *42*(3), 514-532. https://doi.org/10.1007/s12103-016-9366-2

Aos, S., Miller, M., & Drake, E. (2006). Evidence-based adult corrections programs: What works and what does not. Olympia, WA: Washington State Institute for Public Policy.

Austin, J. (2001). Prisoner reentry: Current trends, practices and issues. *Crime and Delinquency, 37,* 314-334. https://doi.org/10.1177/0011128701047003002

Austin, P. C. (2007). Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. *The Journal of Thoracic and Cardiovascular Surgery*, *134*(5), 1128-1135. https://doi.org/10.1016/j.jtcvs.2007.07.021

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine, 28*(25), 3083-3107. https://doi.org/10.1002/sim.3697

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399-424. https://doi.org/10.1080/00273171.2011.568786.

Austin, P. C. (2014). The use of propensity score methods with survival or time-to-event outcomes: Reporting measures of effect similar to those used in randomized experiments. *Statistics in Medicine, 33*(7), 1242-1258. https://doi.org/10.1002/sim.5984

Austin, P. C., & Cafri, G. (2020). Variance estimation when using propensity-score matching with replacement with survival or time-to-event outcomes. *Statistics in Medicine, 39*(11), 1623-1640. https://doi.org/10.1002/sim.8502

Austin, P. C., & Schuster, T. (2016). The performance of different propensity score methods for estimating absolute effects of treatments on survival outcomes: A simulation study. *Statistical Methods in Medical Research, 25*(5), 2214-2237. https://doi.org/10.1177/0962280213519716

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine, 34*(28), 3661-3679. https://doi.org/10.1002/sim.6607

Austin, P. C., Manca, A., Zwarenstein, M., Juurlink, D. N., & Stanbrook, M. B. (2010). A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: A review of trials published in leading medical journals. *Journal of Clinical Epidemiology, 63*(2), 142-153. https://doi.org/10.1016/j.jclinepi.2009.06.002

Baker, V., Johnson, D., & Oluonye, S. (2013). Resettle: A significant new step in an emerging pathway that manages risk and addresses need in high-risk personality disordered offenders on their release into the community. *Psychology, Crime & Law, 19*(5-6), 449-460. https://doi.org/10.1080/1068316X.2013.758976

Balusek, K. (2014). Offender reentry and cognitive intervention: Propensity score matching utility for outcome assessment. LFB Scholarly Publishing LLC. *ProQuest Ebook Central*, https://ebookcentral.proquest.com/lib/bham/detail.action?docID=1756074.

Barkham, M., Bewick, B., Mullin, T., Gilbody, S., Connell, J., Cahill, J., Mellor-Clark, J., Richards, D., Unsworth, G. & Evans, C. (2012). The CORE-10: A short measure of psychological distress for routine use in the psychological therapies. *Counselling and Psychotherapy Research*, 1–11. http://doi.org/10.1080/14733145.2012.729069

Barkham, M., Culverwell, A., Spindler, K., & Twigg, E. (2005). The CORE-OM in an older adult population: Psychometric status, acceptability, and feasibility. *Aging & Mental Health, 9*(3), 235–245. https://doi.org/10.1080/13607860500090052

Barkham, M., Evans, C., Margison, F., & McGrath, G. (1998). The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health, 7*(1), 35–47.

Barkham, M., Gilbert, N., Connell, J., Marshall, C., & Twigg, E. (2005). Suitability and utility of the CORE–OM and CORE–A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *British Journal of Psychiatry, 186*(3), 239–246. https://doi.org/10.1192/bjp.186.3.239

Barkham, M., Hardy, G. E., & Startup, M. (1996). The development of the IIP-32: A short version of the Inventory of Interpersonal Problems. *British Journal of Clinical Psychology, 35*, 21–35. https://doi.org/10.1111/j.2044-8260.1996.tb01159.x

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K., & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE–OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69*(2), 184–196. https://doi.org/10.1037/0022-006X.69.2.184

Barkham, M., Mellor-Clark, J., Connell, J., & Cahill, J. (2006). A core approach to practice-based evidence: A brief history of the origins and applications of the CORE-OM and CORE System. *Counselling and Psychotherapy Research, 6*(1), 3–15. https://doi.org/10.1080/14733140600581218

Bateman, A. W., Gunderson, J., & Mulder, R. (2015). Treatment of personality disorder. *The Lancet, 385*(9969), 735-743. https://doi.org/10.1016/S0140-6736(14)61394-5

Batson, S., Greenall, G., & Hudson, P. (2016). Review of the Reporting of Survival Analyses within Randomised Controlled Trials and the Implications for Meta-Analysis. *PloS One, 11*(5), e0154870. https://doi.org/10.1371/journal.pone.0154870

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*, 893–897. https://doi.org/10.1037/0022-006X.56.6.893

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

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Belitser, S. V., Martens, E. P., Pestman, W. R., Groenwold, R. H., De Boer, A., & Klungel, O. H. (2011). Measuring balance and model selection in propensity score methods. *Pharmacoepidemiology and Drug Safety, 20*(11), 1115-1129. https://doi.org/10.1002/pds.2188

Bell, L., Rogers, R. D., & Coid, J. (2006). Gender differences in DSPD offenders. In *annual conference on dangerous and severe personality disorders, held at Leeds, UK*.

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2018). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research.* https://doi.org/10.1177/0049124118799372

Berghuis, M. (2018). Reentry programs for adult male offender recidivism and reintegration: A systematic review and meta-analysis. *International Journal of Offender Therapy and Comparative Criminology, 62*(14), 4655-4676. https://doi.org/10.1177/0306624X18778448

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 12: survival analysis. *Critical Care, 8*(5), 1-6. https://doi.org/10.1186/cc2955

Blumenthal, S., Craissati, J., & Minchin, L. (2009). The development of a specialist hostel for the community management of personality disordered offenders. *Criminal Behaviour and Mental Health, 19,* 43–53. https://doi.org/10.1002/cbm.720

Bonell, C. P., Hargreaves, J., Cousens, S., Ross, D., Hayes, R., Petticrew, M., & Kirkwood, B. R. (2011). Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions. *Journal of Epidemiology & Community Health, 65*(7), 582-587. http://dx.doi.org/10.1136/jech.2008.082602

Bonta, J., Blais, J., & Wilson, H. A. (2014). A theoretically informed meta-analysis of the risk for general and violent recidivism for mentally disordered offenders. *Aggression and Violent Behavior, 19*(3), 278-287. https://doi.org/10.1016/j.avb.2014.04.014

Bonta, J., & Andrews, D. A. (2016). *The Psychology of Criminal Conduct*. Routledge.

Bouffard, J. A., & Bergeron, L. E. (2006). Reentry works: The implementation and effectiveness of a serious and violent offender reentry initiative. *Journal of Offender Rehabilitation, 44*(2-3), 1-29. https://doi.org/10.1300/J076v44n02_01

Bowlby, J. (1977). The making and breaking of affectional bonds: I. Aetiology and psychopathology in the light of attachment theory. *The British Journal of Psychiatry, 130*(3), 201–210. https://doi.org/10.1192/bjp.130.3.201

Bowlby, J. (1979). The Bowlby-Ainsworth attachment theory. *Behavioral and Brain Sciences, 2*(4), 637-638. https://doi.org/10.1017/S0140525X00064955

Brader, C. (2023). Personality disorders in prison and probation: Are specialist units working? *UK Parliament*. https://lordslibrary.parliament.uk/personality-disorders-in-prison-and-probation-are-specialist-units-working/

Braga, A. A., Piehl, A. M., & Hureau, D. (2009). Controlling violent offenders released to the community: An evaluation of the Boston Reentry Initiative. *Journal of Research in Crime and Delinquency, 46*(4), 411-436. https://doi.org/10.1177/0022427809341935

Bruce, M., Horgan, H., Kerr, R., Cullen, A., & Russell, S. (2017). Psychologically informed practice (PIP) for staff working with offenders with personality disorder: A pragmatic exploratory trial in approved premises. *Criminal Behaviour and Mental Health, 27*: 290-302. https://doi.org/10.1002/cbm.1994

Bruffaerts, R., Demyttenaere, K., Borges, G., Haro, J. M., Chiu, W. T., Hwang, I., Karam, E. G., Kessler, R. C., Sampson, N., Alonso, J., Andrade, L. H., Angermeyer, M., Benjet, C., Bromet, E., de Girolamo, G., de Graaf, R., Florescu, S., Gureje, O., Horiguchi, I., … Nock, M. K. (2010). Childhood adversities as risk factors for onset and persistence of suicidal behaviour. *British Journal of Psychiatry, 197*(1), 20–27. https://doi.org/10.1192/bjp.bp.109.074716

Byrne, G., Hammond, S., & Moore, P. (2010). CORE: The core issues surrounding its psychometric properties: A quantitative analysis of its use in an Irish population of childhood abuse victims. *The British Journal of Forensic Practice, 12*(2), 32–42. https://doi.org/10.5042/bjfp.2010.0184

Cahill, J., Barkham, M., Stiles, W. B., Twigg, E., Hardy, G. E., Rees, A., & Evans, C. (2006). Convergent validity of the CORE measures with measures of depression for clients in

cognitive therapy for depression. *Journal of Counselling Psychology, 53*(2), 253–259. https://doi.org/10.1037/0022-0167.53.2.253

Caliendo, M. & Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys,* 22(1), pp.31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, *36*(6), 567-590. https://doi.org/10.1177/0093854809333610

Carson, E. A. (2020). Prisoners in 2018. US Department of Justice. http://www.antoniocasella.eu/nume/Carson_bjs_apr20.pdf

Cavelti, M., Lerch, S., Ghinea, D., Fischer-Waldschmidt, G., Resch, F., Koenig, J., & Kaess, M. (2021). Heterogeneity of borderline personality disorder symptoms in help-seeking adolescents. *Borderline Personality Disorder and Emotion Dysregulation*, *8*(1), 1-14. https://doi.org/10.1186/s40479-021-00147-9

Chai-Adisaksopha, C., Iorio, A., Hillis, C., Lim, W., & Crowther, M. (2016). A systematic review of using and reporting survival analyses in acute lymphoblastic leukemia literature. *BMC Hematology*, *16*(1), 1-8. https://doi.org/10.1186/s12878-016-0055-7

Clark, V. A. (2015). Making the most of second chances: An evaluation of Minnesota's high-risk revocation reduction reentry program. *Journal of Experimental Criminology*, *11*(2), 193-215. https://doi.org/10.1007/s11292-014-9216-5

Coaley, K. (2010). Introduction: Foundations of psychological assessment. In An Introduction to Psychological Assessment and Psychometrics (pp. 1–22). https://doi.org/10.4135/9781446221556

Coid, J. (1998). Axis II disorders and motivation for serious criminal behaviour, in *Psychopathology and Violent Crime* (ed. A. E. Skodol), American Psychiatric Association, Washington DC, pp. 53-96.

Coid, J., Yang, M., Tyrer, P., Roberts, A. and Ullrich, S. (2006). Prevalence and correlates of personality disorder in Great Britain. *British Journal of Psychiatry, 188*(5), 423-431. https://doi.org/10.1192/bjp.188.5.423

Coid, J., Yang, M., Ullrich, S., Zhang, T., Roberts, A., Roberts, C., Rogers, R., & Farrington, D. (2007). Predicting and understanding risk of re-offending: The prisoner cohort study. *Research Summary 6*(1), 1-9. https://www.crim.cam.ac.uk/sites/www.crim.cam.ac.uk/files/priscomoj.pdf

Collison, K. L., & Lynam, D. R. (2021). Personality disorders as predictors of intimate partner violence: A meta-analysis. *Clinical psychology review*, *88*, 102047. https://doi.org/10.1016/j.cpr.2021.102047

Connell, C., Furtado, V., McKay, E. A., & Singh, S. P. (2018). What influences social outcomes among offenders with personality disorder: A systematic review. *Criminal Behaviour and Mental Health*, *28*(5), 390-396. https://doi.org/10.1002/cbm.2082

Connell, J., Barkham, M., Stiles, W. B., Twigg, E., Singleton, N., Evans, O., & Miles, J. N. V. (2007). Distribution of CORE–OM scores in a general population, clinical cut-off

points and comparison with the CIS–R. *British Journal of Psychiatry, 190*(01), 69–74. https://doi.org/10.1192/bjp.bp.105.017657

Core System Group (1998). CORE System (Information Management) Handbook. Leeds, Core System Group.

Craig, J. M., & Zettler, H. R. (2021). Are the effects of adverse childhood experiences on violent recidivism offense-specific? *Youth Violence and Juvenile Justice, 19*(1), 27-44. https://doi.org/10.1177/1541204020939638

Craig, J. M., Piquero, A. R., Farrington, D. P., & Ttofi, M. M. (2017). A little early risk goes a long bad way: Adverse childhood experiences and life-course offending in the Cambridge study. *Journal of Criminal Justice, 53*, 34-45. https://doi.org/10.1016/j.jcrimjus.2017.09.005

Craissati, J., Ramsden, J., Ryan, S., Webster, N., & West, L. (2021). Intensive intervention and risk management services (IIRMS) three years on: What we need to do better in the Offender Personality Disorder Pathway. *The Journal of Forensic* Practice, *23*(4), 385-396. https://doi.org/10.1108/JFP-08-2021-0042

Craissati, J., Webb, L., & Keen, S. (2008). The relationship between developmental variables, personality disorder and risk in sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 20*(2), 119-138. https://doi.org/10.1177/1079063208317464

Critical Appraisal Skills Programme, UK (2018). CASP Checklists. Retrieved 20th January, 2020, from https://casp-uk.net/casp-tools-checklists/

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Davison, S., & Janca, A. (2012). Personality disorder and criminal behaviour: What is the nature of the relationship? *Current Opinion in Psychiatry*, *25*(1), 39-45. https://doi.org/10.1097/YCO.0b013e32834d18f0

Deci, E. L. & Ryan, R. M. (2000). The ‘‘what’’ and ‘‘why’’ of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227–268. https://doi.org/10.1207/S15327965PLI1104_01

DeLisi, M., Drury, A. J., & Elbert, M. J. (2019). The etiology of antisocial personality disorder: The differential roles of adverse childhood experiences and childhood psychopathology. *Comprehensive Psychiatry*, *92*, 1-6. https://doi.org/10.1016/j.comppsych.2019.04.001

Dennis, J. A., Khan, O., Ferriter, M., Huband, N., Powney, M. J., & Duggan, C. (2012). Psychological interventions for adults who have sexually offended or are at risk of offending. *Cochrane Database of Systematic Reviews*, (12). https://doi.org/10.1002/14651858.CD007507.pub2

Derogatis, L. R. (1993). SCL-90-R: Administration, scoring and procedures: Manual. Towson, MD: Clinical Psychometric Research.

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13*(3), 595–605. https://doi.org/10.1017/S0033291700048017

Desai, R. J., Rothman, K. J., Bateman, B. T., Hernandez-Diaz, S., & Huybrechts, K. F. (2017). A propensity score based fine stratification approach for confounding adjustment when exposure is infrequent. *Epidemiology*, *28*(2), 249-257. https://doi.org/10.1097/EDE.0000000000000595

Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics, 95*(3), 932-945. https://doi.org/10.1162/REST_a_00318

Doherty, F. (2016). Obey all laws and be good: Probation and the meaning of recidivism. *Georgetown Law Journal, 104*(2), 291–355.

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). HCR20V3: Assessing risk for violence – User guide. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Doyal, L., & Gough, I. (1984). A theory of human needs. *Critical Social Policy, 4*(10), 6-38. https://doi.org/10.1177/026101838400401002

Duggan, C. (2009). A treatment guideline for people with antisocial personality disorder: Overcoming attitudinal barriers and evidential limitations. *Criminal Behaviour & Mental Health*, *19*, 219-223. https://doi.org/10.1002/cbm.726

Duncan, G. (2023, September 4). Prison Mental Health Services in Prison, 2023. Centre for Mental Health. https://www.centreformentalhealth.org.uk/publications/prison-mental-health-services-england-2023

Durose, M. R., Cooper, A. D., & Snyder, H. N. (2014). Recidivism of Prisoners in 30 States: Patterns from 2005-2010. *Bureau of Justice Statistics*. https://bjs.ojp.gov/content/pub/pdf/rprts05p0510.pdf

Duwe, G., & McNeeley, S. (2021). The effects of intensive post-release correctional supervision on recidivism: A natural experiment. *Criminal Justice Police Review, 32*(7), 740-763. https://doi.org/10.1177/0887403421998430

Elfström, M. L., Evans, C., Lundgren, J., Johansson, B., Hakeberg, M., & Carlsson, S. G. (2013). Validation of the Swedish version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM). *Clinical Psychology & Psychotherapy, 20*(5), 447-455. https://doi.org/10.1002/cpp.1788

Engelen, P-J., Lander, M. W., & van Essen, M. (2016). What determines crime rates? An empirical test of integrated economic and sociological theories of criminal behaviour. *The Social Science Journal, 53*(2), 247-262. https://doi.org/10.1016/j.soscij.2015.09.001

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE–OM. *British Journal of Psychiatry, 180*(1), 51–60. https://doi.org/10.1192/bjp.180.1.51

Evans, C., Connell, J., Barkham, M., Marshall, C., & Mellor-Clark, J. (2003). Practice-based evidence: Benchmarking NHS primary care counselling services at national and local levels. *Clinical Psychology & Psychotherapy, 10*(6), 374–388. https://doi.org/10.1002/cpp.384

Evans, J., Mellor-Clark, F., Margison, M., Barkham, K., Audin, J., Connell, G., McGrath, C. (2000). CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health, 9*(3), 247-255. https://doi.org/10.1080/jmh.9.3.247.255

Falshaw, L., Friendship, C., Travers, R., & Nugent, F. (2003). Searching for 'what works': An evaluation of cognitive skills programmes. Research findings no. 206. London, UK: Home Office Research. *Development and Statistics Directorate*.

Farrall, S., & Calverley, A. (2005). *Understanding desistance from crime*. McGraw-Hill Education (UK).

Fazel, S., & Danesh, J. (2002). Serious mental disorder in 23,000 prisoners: A systematic review of 62 surveys. *The Lancet, 359*, 545-550. https://doi.org/10.1016/S0140-6736(02)07740-1

Fazel, S., & Seewald, K. (2012). Severe mental illness in 33,588 prisoners worldwide: Systematic review and meta-regression analysis. *British Journal of Psychiatry, 200,* 364–73. https://doi.org/10.1192/bjp.bp.111.096370

Fazel, S., & Wolf, A. (2015). A Systematic Review of Criminal Recidivism Rates Worldwide: Current Difficulties and Recommendations for Best Practice. *PLoS One, 10*(6). https://doi.org/10.1371/journal.pone.0130390

Fink, A. (2010). Survey research methods. *International Encyclopedia of Education, 3*, 152-160. https://doi.org/10.1016/B978-0-08-044894-7.00296-7

Fitzpatrick, R., Chambers, J., Burns, T., Doll, H., Fazel, S., Jenkinson, C., Kaur, A., Knapp, M., Sutton, L., & Yiend, J. (2010). A systematic review of outcome measures used in forensic mental health research with consensus panel opinion. *Health Technology Assessment, 14*(18), 1-94. https://doi.org/10.3310/hta14180

Fleiss, J. L. (2011). Design and Analysis of Clinical Experiments. John Wiley & Sons. https://doi.org/10.1002/9781118032923

Flynn, R. (2012). Survival Analysis. *Journal of Clinical Nursing, 21,* 2789–2797. https://doi.org/10.1111/j.1365-2702.2011.04023.x

Fonagy, P., Matthews, R. & Pilling, S. (2004) The mental health outcomes measurement initiative: Report from the chair of the outcomes reference group. London: National Collaborating Centre for Mental Health, Department of Health.

Fox, B. H., Perez, N., Cass, E., Baglivio, M. T., & Epps, N. (2015). Trauma changes everything: Examining the relationship between adverse childhood experiences and serious, violent and chronic juvenile offenders. *Child Abuse & Neglect, 46,* 163-173. https://doi.org/10.1016/j.chiabu.2015.01.011

Francis, B., Soothill, K. & Humphreys, L. (2007). *Development of a reoffending measure using the Police National Computer database.* Centre for Applied Statistics, Lancaster University, p.1-45. https://eprints.lancs.ac.uk/id/eprint/50057/

Fridell, M., Hesse, M., Jaeger, M. M., & Kuhlhorn, E. (2008). Antisocial personality disorder as a predictor of criminal behaviour in a longitudinal study of a cohort of abusers of several classes of drugs: Relation to type of substance and type of crime. *Addictive Behaviours, 33,* 799-811. https://doi.org/10.1016/j.addbeh.2008.01.001

Friendship, C., Mann, R. E., & Beech, A. R. (2003). Evaluation of a national prison-based treatment program for sexual offenders in England and Wales. *Journal of Interpersonal Violence, 18*(7), 744-759. https://doi.org/10.1177/0886260503253236

Froyd J. E., Lambert M. J., & Froyd J. D. (1996). A Review of Practices of Psychotherapy Outcome Measurement. Oxfordshire, Routledge. https://doi.org/10.1080/09638239650037144

Gayat, E., Pirracchio R., Resche-Rigon, M., Mebazaa, A., Mary, J-Y., & Porcher, R. (2010) Propensity scores in intensive care and anaesthesiology literature: A systematic review. *Intensive Care Medicine, 36*(12), 1993-2003. https://doi.org/10.1007/s00134-010-1991-5

Gayat, E., Resche-Rigon, M., Mary, J. Y., & Porcher, R. (2012). Propensity score applied to survival data analysis through proportional hazards models: A Monte Carlo study. *Pharmaceutical Statistics, 11*(3), 222-229. https://doi.org/10.1002/pst.537

Gibson, C. L., Swatt, M. L., Miller, J. M., Jennings, W. G., & Gover, A. R. (2012). The causal relationship between gang joining and violent victimization: A critical review and directions for future research. *Journal of Criminal Justice, 40*(6), 490–501. https://doi.org/10.1016/j.jcrimjus.2012.08.001

Giromini, L., Young, G., & Sellbom, M. (2022) Assessing negative response bias using self report measures: New articles, new issues. *Psychological Injury and Law, 15*, 1–21. https://doi.org/10.1007/s12207-022-09444-2

Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine, 9*(1), 139–145. https://doi.org/10.1017/S0033291700021644

Golinelli, D., Ridgeway, G., Rhoades, H., Tucker, J., & Wenzel, S. (2012). Bias and variance trade-offs when combining propensity score weighting and regression: With an application to HIV status and homeless men. *Health Services and Outcomes Research Methodology, 12*(2), 104-118. https://doi.org/10.1007/s10742-012-0090-1

Gottfried, E. D., & Christopher, S. C. (2017). Mental disorders among criminal offenders: A review of the literature. *Journal of Correctional Health Care, 23*(3), 336-346. https://doi.org/10.1177/1078345817716180

Grace, S., Lloyd, C., & Smith, L. (1992). *Rape: From recording to conviction*. Research and Planning Unit Paper No 71. London: Home Office.

Greenberg, D. F. (1999). Punishment, division of labor, and social solidarity. *The Criminology of Criminal Law, 8*.

Greifer, N. (2022a, November 15). Assessing Balance. R Project. https://cran.r-project.org/web/packages/MatchIt/vignettes/assessing-balance.html.

Greifer, N. (2022b, November 15). *Matching Methods.* R Project. https://cran.r-project.org/web/packages/MatchIt/vignettes/matching-methods.html.

Greifer, N., & Stuart, E. A. (2021). Choosing the estimand when matching or weighting in observational studies. https://doi.org/10.48550/arXiv.2106.10577

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research, 2*(3), 109. https://doi.org/10.4103/2229-3485.83221

Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics, 15*(3), 609-627. https://doi.org/10.1198/106186006X137047

Hanson, R. K., & Bussiere, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, *66*(2), 348.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*(1), 119-136.

Hanson, R. K., Broom, I., & Stephenson, M. (2004). Evaluating community sex offender treatment programs: A 12-year follow-up of 724 offenders. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement, 36*(2), 87-96. https://doi.org/10.1037/h0087220

Harrison, L., & Alves-Costa, F. (2022). Predictors associated with readmission in low secure forensic settings in the UK: An exploratory investigation and survival analysis. *The Journal of Forensic Psychiatry & Psychology, 33*(1), 178-196. https://doi.org/10.1080/14789949.2021.2007981

Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). The Risk for Sexual Violence Protocol (RSVP): Structured professional guidelines for assessing risk of sexual violence. Burnaby, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University.

Hayes, V., Morris, J., Wolfe, C., & Morgan, M. (1995). The SF-36 Health Survey Questionnaire: Is it suitable for use with older adults? *Age and Ageing, 24*(2), 120-125. https://doi.org/10.1093/ageing/24.2.120

Hedderman, C., & Sugg, D. (1996). Does Treating Sex Offenders Reduce Re-Offending? Research findings no. 45. London, UK: Home Office Research. *Development and Statistics Directorate.*

Hernandez-Avila, C. A., Burleson, J. A., Poling, J., Tennen, H., Rounsaville, B. J., & Kranzler, H. R. (2000). Personality and substance use disorders as predictors of criminality. *Comprehensive Psychiatry*, *41*(4), 276-283. https://doi.org/10.1053/comp.2000.7423

Home Office (2018, November 20). Home Office counting rules for recorded crime. Data.gov.uk. https://www.data.gov.uk/dataset/695f6775-3e51-4dd4-911a-19575638384c/home-office-counting-rules-for-recorded-crime

HM Prison and Probation Service (2022). Offender management statistics quarterly: October to December 2021 and annual 2021 - GOV.UK (www.gov.uk)

HM Prison and Probation Service (2022). *Risk of Serious Harm Guidance 2020*. Public Protection Group, v.2. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1060610/Risk_of_Serious_Harm_Guidance_March_2022.pdf

HM Prison and Probation Service and National Health Service (2020). Practitioner Guide: Working with people in the Criminal Justice System showing Personality Difficulties. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachme

nt_data/file/1035881/6.5151_HMPPS_Working_with_Offenders_with_Personality_Disorder_accessible_version_.pdf

HM Prison and Probation Service and National Health Service (2021). Principles and practice: Guidance for implementing the Intensive Intervention and Risk Management Service (IIRMS) Specification. https://www.justiceinspectorates.gov.uk/hmiprobation/wp-content/uploads/sites/5/2022/01/IIRMS-guidance-final-Nov-21.pdf

HM Prison and Probation Service in Wales (2018). *Wales Integrated Serious and Dangerous Offender Management Operating Model*, v.8.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199-236. https://doi.org/10.1093/pan/mpl013

Hoffman, P. B., & Stone-Meierhoefer, B. (1980). Reporting recidivism rates: The criterion and follow-up issues. *Journal of Criminal Justice*, *8*(1), 53-60. https://doi.org/10.1016/0047-2352(80)90059-8

Home Office & Department of Health (1999). Managing dangerous people with severe personality disorder. Proposals for Policy Development. *London: Stationery Office*. https://doi.org/10.1080/09627259908552910

Howard, P., Francis, B., Soothill, K., & Humphreys L. (2009). OGRS 3: The revised Offender Group Reconviction Scale. https://eprints.lancs.ac.uk/id/eprint/49988/1/ogrs3.pdf

Howard, R. C., Huband, N., Duggan, C., & Mannion, A. (2008). Exploring the link between personality disorder and criminality in a community sample. *Journal of Personality Disorders*, *22*(6), 589-603. https://doi.org/10.1521/pedi.2008.22.6.589

Integrated Response Integrated Service (2020). IRiS Offender Management Multi-Agency Agreement (v.5).

Jacobs, D., & Kleban, R. (2003). Political institutions, minorities, and punishment: A pooled cross-national analysis of imprisonment rates. *Social Forces*, *82*(2), 725-755. https://doi.org/10.1353/sof.2004.0013

Jacobson, N. S., & Truax, P. (1991). Clinical Significance: A Statistical Approach to Denning Meaningful Change in Psychotherapy Research, 8, 12-19.

Jewell, A., Cocks, C., Cullend, A. E., Fahya, T., & Dean, K. (2018). Predicting time to recall in patients conditionally released from a secure forensic hospital: A survival analysis. *European Psychiatry*, *49*, 1–8. https://doi.org/10.1016/j.eurpsy.2017.11.005

Jolliffe, D., Cattell, J., Raza, A., & Minoudis, P. (2017). Evaluating the impact of the London Pathway Project. *Criminal Behaviour and Mental Health*, *27*(3), 238–253. https://doi.org/10.1002/cbm.2041

Joseph, N., & Benefield, N. (2012). A joint Offender Personality Disorder Pathway strategy: An outline summary. *Criminal Behaviour and Mental Health*, *22*(3), 210–217. https://doi.org/10.1002/cbm.1835

Jung, S., Wielinga, F., & Ennis, L. (2018) Can we keep it simple? Using the BARR-2002R with a community-based sex offender sample. *Journal of Sexual Aggression, 24*(1), 25-36, https://doi.org/10.1080/13552600.2017.1388643

Kainz, K., Greifer, N., Givens, A., Swietek, K., Lombardi, B. M., Zietz, S., & Kohn, J. L. (2017). Improving causal inference: Recommendations for covariate selection and balance in propensity score methods. *Journal of the Society for Social Work and Research, 8*(2), 279-303. https://doi.org/10.1086/691464

Kerr, T., Fairbairn, N., Tyndall, M., Marsh, D., Li, K., Montaner, J., & Wood, E. (2007). Predictors of non-fatal overdose among a cohort of polysubstance-using injection drug users. *Drug and alcohol dependence, 87*(1), 39-45. https://doi.org/10.1016/j.drugalcdep.2006.07.009

Kim, R. H., & Clark, D. (2013). The effect of prison-based college education programs on recidivism: Propensity score matching approach. *Journal of Criminal Justice, 41*(3), 196-204. https://doi.org/10.1016/j.jcrimjus.2013.03.001

Kinner, S. A., & Wang, E. A. (2014). The case for improving the health of ex-prisoners. *American Journal of Public Health, 104*(8), 1352-1355. https://doi.org/10.2105/AJPH.2014.301883

Kinner, S. A., Forsyth, S., & Williams, G. M. (2013). Systematic review of record linkage studies of mortality in ex-prisoners: why (good) methods matter. *Addiction. 108*(1), 38-49. https://doi.org/10.1111/add.12010

Kinner, S. A., Milloy, M. J., Wood, E., Qi, J., Zhang, R., & Kerr, T. (2012). Incidence and risk factors for non-fatal overdose among a cohort of recently incarcerated illicit drug users. *Addictive Behaviors, 37*(6), 691-696. https://doi.org/10.1016/j.addbeh.2012.01.019

Kirk, D. S., & Wakefield, S. (2018). Collateral consequences of punishment: A critical review and path forward. *Annual Review of Criminology, 1*(1), 171–194. https://doi.org/10.1146/annurev-criminol-032317-092045

Kline, P. (1986). A handbook of test construction: Introduction to psychometric design. London: Methuen.

Knight, C. (2019). Trauma informed practice and care: Implications for field instruction. *Clinical Social Work Journal, 47*, 79–89. https://doi.org/10.1007/s10615-018-0661-x

Kotterbova, E., & Lad, S. (2022). Predictors of engagement in female offenders accessing mental health treatment requirements. *The Journal of Forensic Psychiatry & Psychology, 33*(1), 53-67. https://doi.org/10.1080/14789949.2021.1884737

Langan, P. A., & Levin, D. J. (2002). Recidivism of prisoners released in 1994. BJS Special Report, NCJ 193427. U.S. Department of Justice: Washington, D.C.

Lattimore, P. K., & Visher, C. A. (2009). The multi-site evaluation of SVORI: Summary and synthesis. Washington, D.C.: Urban Institute. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=877cdfb8a35c173910520b4f3a7cad32179de57a

Leach, C., Lucock, M., Barkham, M., Noble, R., Clarke, L., & Iveson, S. (2005). Assessing risk and emotional disturbance using the CORE–OM and HoNOS outcome measures

at the interface between primary and secondary mental healthcare. *Psychiatric Bulletin, 29*(11), 419–422. https://doi.org/10.1192/pb.29.11.419

Leach, C., Lucock, M., Barkham, M., Stiles, W. B., Noble, R., & Iveson, S. (2006). Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice. *British Journal of Clinical Psychology, 45*, 153-166. https://doi.org/10.1348/014466505X35335

Lewis, G., Pelosi, A. J., Araya, R., & Dunn, G. (1992) Measuring psychiatric disorder in the community: A standardized assessment for use by lay interviewers. *Psychological Medicine, 22*, 465-486. https://doi.org/10.1017/S0033291700030415

Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science, 3,* 297-320. https://doi.org/10.1146/annurev.lawsocsci.3.081806.112833

Logan, C., & Johnstone, L. (2010). Personality disorder and violence: Making the link through risk formulation. *Journal of Personality Disorders*, *24*(5), 610-633. https://doi.org/10.1521/pedi.2010.24.5.610

Losel, F. (2001). Evaluating the effectiveness of correctional programs: Bridging the gap between research and practice. In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), Offender rehabilitation in practice (pp. 67-92). Chichester, UK: Wiley.

Lovell, L., & Hardy, G. (2014). Having a diagnosis of borderline personality disorder in a forensic setting: A qualitative exploration. *The Journal of Forensic Practice, 16*(3), 228–240. https://doi.org/10.1108/JFP-01-2014-0003

Macinnes, M., Macpherson, G., Austin, J., & Schwannauer, M. (2016). Examining the effect of childhood trauma on psychological distress, risk of violence and engagement, in forensic mental health. *Psychiatric Research, 246,* 314-320. http://dx.doi.org/10.1016/j.psychres.2016.09.054

Mackenzie, D. L. (2006). What works in corrections: Reducing the criminal activities of offenders and delinquents. New York: Cambridge University Press.

Maguire, M. (2013). The resettlement of ex-prisoners. In *Handbook of Probation* (pp. 420-446). Willan.

Manchak, S. M., Skeem, J. L., & Douglas, K. S. (2008). Utility of the Revised Level of Service Inventory (LSI-R) in predicting recidivism after long-term incarceration. *Law and Human Behavior, 32*, 477-488. https://doi.org/10.1007/s10979-007-9118-4

Marshall, W. L., & Marshall, L. E. (2007). The utility of the random controlled trial for evaluating sexual offender treatment: The gold standard or an inappropriate strategy? *Sexual Abuse: A Journal of Research and Treatment*, *19*(2), 175-191. https://doi.org/10.1007/s11194-007-9046-3

Maruna, S. (2001) *Making Good: How Ex-Convicts Reform and Rebuild Their Lives*. American Psychological Association Books, Washington DC.

Maruna, S., & Mann, R. (2019). Reconciling 'desistance' and 'what works'. *Academic Insights, 1*, 3-10. https://inspectorates-staging.bang-on.net/hmiprobation/wp-content/uploads/sites/5/2019/02/Academic-Insights-Maruna-and-Mann-Feb-19-final.pdf

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*, 370-396.

Matz, A. K. (2018). Community Corrections and the Health of Criminal Justice Populations. *Journal of Health & Human Services Administration, 41*(3), 348-383. https://www.jstor.org/stable/26974604

Mawby, Z., Newman, A., & Wilkinson-Tough, M. (2020). Offender personality disorder pathway screening tools evaluation. *Journal of Forensic Practice, 22*(3), 199-211, https://doi.org/10.1108/JFP-09-2019-0043

McCloskey, H. (2001). Evaluation of the CORE outcome measure in a therapeutic forensic setting. *The British Journal of Forensic Practice, 3*(2), 22-28. https://doi.org/10.1108/14636646200100011

McDonald, J. D. (2008). Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire, 1*(1), 1-19. https://www.nottingham.ac.uk/sociology/documents/enquire/volume-1-issue-1-dodorico-mcdonald.pdf

McIntosh, L. G., Slesser, M., O'Rourke, S., & Thompson, L. D. G. (2018). On the road to recovery psychological therapy versus treatment as usual for forensic mental health patients: Study protocol for a randomized controlled feasibility trial. *Pilot and Feasibility Studies, 4*, 1-11. https://doi.org/10.1186/s40814-018-0319-z

McMurran, M., & Howard, R. (Eds.). (2009). Personality, Personality Disorder and Violence: An evidence-based approach. John Wiley & Sons.

McMurran, M., & Theodosi, E. (2007). Is treatment non-completion associated with increased reconviction over no treatment? *Psychology, Crime & Law*, *13*(4), 333-343. https://doi.org/10.1080/10683160601060374

Mellor-Clark, J., Barkham, M., Evans, C., Margison, F., McGrath, G. & Connell, J. (1997). Talking effectiveness: Providers' and purchasers' perspectives on outcomes. PTRC Memo 335, Psychological Therapies Research Centre, University of Leeds.

Merrall, E. L., Kariminia, A., Binswanger, I. A., Hobbs, M. S., Farrell, M., Marsden, J., Hutchinson, S. J., & Bird, S. M. (2010). Meta-analysis of drug-related deaths soon after release from prison. *Addiction, 105*(9), 1545-1554. https://doi.org/10.1111/j.1360-0443.2010.02990.x

Ming, K., & Rosenbaum, P. R. (2000). Substantial gains in bias reduction from matching with a variable number of controls. *Biometrics, 56*(1), 118–124. https://doi.org/10.1111/j.0006-341X.2000.00118.x

Ministry of Justice (2011). Working with Personality Disordered Offenders: A Practitioners Guide. London: National Offender Management Service and Department of Health. https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2015/10/work-offndrs-persnlty-disorder-oct15.pdf

Ministry of Justice (2012). Probation trust unit costs: Financial year 2011-12 (revised). *Ministry of Justice Information Release*. https://publications.parliament.uk/pa/cm201719/cmselect/cmjust/482/48208.htm#footnote-145

Ministry of Justice (2015). Integrated Offender Management: Key Principles. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/406865/HO_IOM_Key_Principles_document_Final.pdf

Ministry of Justice (2019). Costs per place and costs per prisoner by individual prison. HM Prison and Probation Service Annual Report and Accounts 2018-19. *Ministry of Justice Information Release.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/841948/costs-per-place-costs-per-prisoner-2018-2019.pdf

Ministry of Justice (2019). Offender Management Statistics Bulletin, England and Wales.

Ministry of Justice (2021). National Statistics: Ethnicity and the criminal justice system, 2020. https://www.gov.uk/government/statistics/ethnicity-and-the-criminal-justice-system-statistics-2020/ethnicity-and-the-criminal-justice-system-2020#offender-management

Ministry of Justice (2022). Offender Management Statistics Bulletin, England and Wales. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1071778/OMSQ_Q4_2021.pdf

Ministry of Justice and HM Prison and Probation Service (2022). Generic Parole Process Policy Framework. https://www.gov.uk/government/publications/generic-parole-process-policy-framework

Minnesota Department of Corrections. (2006). Final Report on the Serious Offender Accountability Restoration (SOAR) Project. St. Paul, MN: Minnesota Department of Corrections. https://www.doc.state.mn.us/publications/documents/ProjectSOAR_FinalReport10-06.pdf

Minnesota Department of Corrections (2011). An Evaluation of the Prisoner Reentry Initiative: Final Report. St. Paul, MN: Minnesota Department of Corrections. https://www.doc.state.mn.us/publications/documents/PRIEvaluationReportFinal.pdf

Minoudis, P., Shaw, J., Bannerman, A., & Craissati, J. (2012). Identifying personality disturbance in a London probation sample. *Probation Journal, 59*(1), 23-38. https://doi.org/10.1177/0264550511429842

Moore, S. (2009). Prison spending outpaces all but Medicaid. *The New York Times, 2.* http://www.nytimes.com/2009/03/03/us/03prison.html

Nadesu, A. (2007). Reconviction patterns of released prisoners: A 36-months follow-up analysis. Wellington, NZ: Department of Corrections. https://www.corrections.govt.nz/__data/assets/pdf_file/0015/10680/reimprisonment-report.pdf

Nathan, R., Centifanti, L., Baker, V., & Hill, J. (2019). A pilot randomised controlled trial of a programme of psychosocial interventions (Resettle) for high-risk personality disordered offenders. *International Journal of Law and Psychiatry, 66*, 1-9. https://doi.org/10.1016/j.ijlp.2019.101463

National Institute for Mental Health in England. (2003). *Personality Disorder: No Longer a Diagnosis of Exclusion: Policy Implementation Guidance for the Development of Services for People with Personality Disorder*. NIMHE.

Nicholaichuk, T. P., Olver, M. E., Gu, D., & Wong, S. C. P. (2014). Age, actuarial risk, and long-term recidivism in a national sample of sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 26,* 406–428. https://doi.org/10.1177/1079063213492340

O'Meara, A., Morgan, L., Godden, S., & Davies, J. (2019). A model of a specialist transitional support and liaison service within the Offender Personality Disorder Pathway in Wales: Learning from a regional pilot service. *Probation Journal, 66*(3), 348-355. https://doi.org/10.1177/0264550519859067

Offender Personality Disorder Pathway (2018). *Guidance for Implementing the Intensive Intervention and Risk Management Service (IIRMS) Specification*: Principles and Practice.

Olmos, A., & Govindasamy, P. (2014). *Randomized experiments vs. Propensity scores matching: A Meta-analysis*. Paper presented at the American Evaluation Association, Denver, CO.

Olmos, A., & Govindasamy, P. (2015). Propensity scores: A practical introduction using R. *Journal of Multi-Disciplinary Evaluation, 11*(25), 68-88. https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/431/414

Parloff, M. (1998). Measuring patient changes in mood, anxiety, and personality disorders: Toward a care battery. Hans H. Strupp, Leonard M. Horowitz, and Michael J. Lambert, (Eds.) Washington, DC: American Psychological Association. *Psychotherapy Research, 8*(3), 356-359. https://doi.org/10.1080/10503309812331332447

Pearce, S., Scott, L., Attwood, G., Saunders, K., Dean, M., De Ridder, R., Galea, D., Konstantinidou, H., & Crawford, M. (2017). Democratic therapeutic community treatment for personality disorder: Randomised controlled trial. *The British Journal of Psychiatry, 210*(2), 149-156. https://doi.org/10.1192/bjp.bp.116.184366

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys, 3,* 96–146. https://doi.org/10.1214/09-SS057

Penney, S. R., Marshall, L., & Simpson, A. I. (2018). A prospective study of pathways to hospital readmission in Canadian forensic psychiatric patients. *The Journal of Forensic Psychiatry & Psychology, 29*(3), 368-386, https://doi.org/10.1080/14789949.2017.1395061

Perry, T., Barkham, M., & Evans, C. (2013). The CORE-OM and CORE-OM (SV) in secure settings: A template analysis of the experiences of male patients and their staff. The *Journal of Forensic Practice, 15*(1), 32–43. https://doi.org/10.1108/14636641311299068

Peters, D. J., Hochstetler, A., DeLisi, M., & Kuo, H. (2015). Parolee recidivism and successful treatment completion: Comparing hazard models across propensity methods. *Journal of Quantitative Criminology, 31*(1), 149–181. https://doi.org/10.1007/s10940-014-9229-2

Petersilia, J. (2011). Beyond the prison bubble. *The Wilson Quarterly (1976-), 35*(1), 50-55. http://archive.wilsonquarterly.com/sites/default/files/articles/WQ_VOL35_W_2011_Article_05_1.pdf

Phelps, M. S. (2017). Mass probation: Toward a more robust theory of state variation in punishment. *Punishment & Society, 19*(1), 53–73. https://doi.org/10.1177/1462474516649174

Phelps, M. S. (2020). Mass probation from micro to macro: Tracing the expansion and consequences of community supervision. *Annual Review of Criminology, 3*, 261–279. https://doi.org/10.1146/annurev-criminol-011419-041352

Polaschek, D. L., & Yesberg, J. A. (2018). High-risk violent prisoners' patterns of change on parole on the DRAOR's dynamic risk and protective factors. *Criminal Justice and Behavior, 45*(3), 340-363. https://doi.org/10.1177/0093854817739928

Polaschek, D. L., Yesberg, J. A., & Chauhan, P. (2018). A year without a conviction: An integrated examination of potential mechanisms for successful reentry in high-risk violent prisoners. *Criminal Justice and Behavior, 45*(4), 425-446. https://doi.org/10.1177/0093854817752757

Pournaghash-Tehrani, S., & Feizabadi, Z. (2009). Predictability of physical and psychological violence by early adverse childhood experiences. *Journal of Family Violence, 24*, 417-422. https://doi.org/10.1007/s10896-009-9245-4

Procter, E. (1996). A five-year outcome evaluation of a community-based treatment programme for convicted sexual offenders run by the probation service. *Journal of Sexual Aggression, 2*(1), 3-16. https://doi.org/10.1080/13552609608413253

Rakis, J. (2005). Improving the employment rates of ex-prisoners under parole. *Federal Probation, 69*(1), 7-12.

Ramsden, J. (2018). "Are you calling me a liar"? Clinical interviewing more for trust than knowledge with high-risk men with antisocial personality disorder. *International Journal of Forensic Mental Health, 17*(4), 351-361. https://doi.org/10.1080/14999013.2018.1505789

Ramsden, J., Hirons, A., Maltman, L., & Mullen, T. (2017). Finding our way: Early learning from the Compass Project, an Intensive Intervention Risk Management service for women. *The Journal of Forensic Psychiatry & Psychology, 28*(2), 257-273. https://doi.org/10.1080/14789949.2016.1237537

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine, 26*(1), 20–36. https://doi.org/10.1002/sim.2739

Ryan, S., Eldridge, A., Duffy, C., Crawley, E., & O'Brien, C. (2022). Resettle Intensive Intervention and Risk Management Service (IIRMS): A pathway to desistance? *The Journal of Forensic Practice, 24*(4), 364-375. https://doi.org/10.1108/JFP-04-2022-0016

Ryan, S., Gordon, F., & Gordon, N. (2019). Relational fragility, development and an emerging self: Service user views of engaging in an IIRMS in the OPD pathway. *Journal of Forensic Practice, 21*(3), 169-179. https://doi.org/10.1108/JFP-02-2019-0007

Sartori, R., & Pasini, M. (2007). Quality and quantity in test validity: How can we be sure that psychological tests measure what they have to? *Quality & Quantity, 41*, 359-374. https://doi.org/10.1007/s11135-006-9006-x

Saunders, D. G. (1991). Procedures for adjusting self-reports of violence for social desirability bias. *Journal of Interpersonal violence*, *6*(3), 336-344. https://doi.org/10.1177/088626091006003006

Seewald, K., Rossegger, A., Gerth, J., Urbaniok, F., Phillips, G., & Endrass, J. (2018). Effectiveness of a risk–need–responsivity-based treatment program for violent and sexual offenders: Results of a retrospective, quasi-experimental study. *Legal and Criminological Psychology, 23*(1), 85–99. https://doi.org/10.1111/lcrp.12122

Segerstrom, S. C., & Smith, G. T. (2019). Personality and coping: Individual differences in responses to emotion. *Annual Review of Psychology, 70*, 651-671. https://doi.org/10.1146/annurev-psych-010418-102917

Seiter, R., & Kadela, K. (2003). Prisoner reentry: What works, what does not, and what is promising. *Crime & Delinquency, 49*, 360-388. https://doi.org/10.1177/0011128703049003002

Shadish, W. R. (2013). Propensity score analysis: Promise, reality and irrational exuberance. *Journal of Experimental Criminology, 9*(2), 129–144. https://doi.org/10.1007/s11292-012-9166-8

Shadish, W. R., & Steiner, P. M. (2010). A primer on propensity score analysis. *Newborn and Infant Nursing Reviews, 10*(1), 19–26. https://doi.org/10.1053/j.nainr.2009.12.010

Shaul, L., Koeter, M. W., & Schippers, G. M. (2016). Brief motivation enhancing intervention to prevent criminal recidivism in substance-abusing offenders under supervision: A randomized trial. *Psychology, Crime & Law*, *22*(9), 903-914. https://doi.org/10.1080/1068316X.2016.1202248

Shuker, R. (2010). Personality disorder: Using therapeutic communities as an integrative approach to address risk. Grendon and the emergence of forensic therapeutic communities: Developments in research and practice, 113-136.

Singleton, N., Gatward, R., & Meltzer, H. (1998). Psychiatric morbidity among prisoners in England and Wales. London: Stationery Office.

Singleton, N., Meltzer, H., Gatward, R., Coid, J., & Deasy, D. (1998). Psychiatric morbidity among prisoners: Summary report.

Skett, S., Goode, I. & Barton, S. (2017). A join NHS and NOMS offender personality disorder pathway strategy: a perspective from 5 years of operation. *Criminal Behaviour and Mental Health, 27*(3), 214-221. https://doi.org/10.1002.cbm.2026

Smith, M. & Smith, P. (2005). Testing people at work: Competencies in psychometric testing. Oxford: BPS Blackwell.

Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T., Yoshinobu, L., Gibb, J., Langelle, C., & Harney, P. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. *Journal of Personality and Social Psychology, 60*(4), 570-585. https://doi.org/10.1037/0022-3514.60.4.570

Social Exclusion Unit (2002). Reducing re-off ending by ex-prisoners. https://www.bristol.ac.uk/poverty/downloads/keyofficialdocuments/Reducing%20Reoffending.pdf

Steiner, P. M., Cook, T. D., & Shadish, W. R. (2011). On the importance of reliable covariate measurement in selection bias adjustments using propensity scores. *Journal of Educational and Behavioral Statistics, 36*(2), 213–236. https://doi.org/10.3102/1076998610375835

Stensrud, R. H., Gilbride, D. D., & Bruinekool, R. M. (2019). The childhood to prison pipeline: Early childhood trauma as reported by a prison population. *Rehabilitation Counselling Bulletin, 62*(4), 195-208. https://doi.org/10.1177/0034355218774844

Stone, M. H. (2007). Violent crimes and their relationship to personality disorders. *Personality and Mental Health,* 1, 138-153. https://doi.org/10.1002/pmh.18

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). Health measurement scales: A practical guide to their development and use. Oxford University Press, USA. https://doi.org/10.1093/med/9780199685219.001.0001

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science: A review journal of the Institute of Mathematical Statistics, 25*(1), 1-21. https://doi.org/10.1214/09-STS313

Sturge, G. (2018, July 23). *UK Prison Population Statistics*. *Briefing Paper Number CBP-04334.* House of Commons Library. https://bulger.co.uk/prison/UKprisonPopulation2018.pdf

Sturge, G. (2022, October 25). *UK Prison Population Statistics. Number 04334*. House of Commons Library. https://researchbriefings.files.parliament.uk/documents/SN04334/SN04334.pdf

Taylor, J. & Trout, S. (2013). Lessons from the front line: Working with offenders with learning disability and personality disorder in a high secure therapeutic community. *Mental Health Review Journal, 18*(1), 44-52. https://doi-org.bham-ezproxy.idm.oclc.org/10.1108/13619321311306974

The Sentencing Project (2019). *Fact Sheet: Trends in U.S. Corrections.* https://www.gnjumc.org/content/uploads/2016/03/Trends-in-Corrections-Fact-sheet.pdf

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research, 46*(1), 90-118. https://doi.org/10.1080/00273171.2011.540475

Travis, J. & Petersilia, J. (2001). Reentry reconsidered: A new look at an old question. *Crime & Delinquency, 47*(3), 291-313. https://doi.org/10.1177/0011128701047003001

Travis, J. (2005). *But they all come back: Facing the challenges of prisoner reentry*. The Urban Institute.

Travis, J., & Visher, C. (2005). Prisoner reentry and the pathways to adulthood: Policy perspectives. O*n your own without a net: The transition to adulthood for vulnerable populations*, 145-177.

Tyrer, P., Nur, U., Crawford, M., Karlsen, S., McLean, C., Rao, B., & Johnson, T. (1985). Social Functioning Questionnaire. *International Journal of Social Psychiatry*. https://doi.org/10.1037/t03074-000

Urbina, S. (2014). Essentials of Psychological Testing. John Wiley & Sons. https://books.google.co.uk/books?hl=en&lr=&id=UnHrAwAAQBAJ&oi=fnd&pg=PR9&dq=psychological+testing&ots=VeBcuH69lu&sig=oCoU40Cq96LHjTQUAA_DMUCG-dY&redir_esc=y#v=onepage&q=psychological%20testing&f=false

US Department of Justice (2021). Probation and Parole in the United States 2020. https://bjs.ojp.gov/content/pub/pdf/ppus20.pdf

Walter, M., Wiesbeck, G. A., Dittmann, V., & Graf, M. (2011). Criminal recidivism in offenders with personality disorders and substance use disorders over 8 years of time at risk. *Psychiatry Research, 186*(2-3), 443-445. https://doi.org/10.1016/j.psychres.2010.08.009

Ward, T., & Stewart, C. (2003). Criminogenic needs and human needs: A theoretical model. *Psychology, Crime & Law*, *9*(2), 125-143. https://doi.org/10.1080/1068316031000116247

Ward, T., & Willis, G. (2010). Ethical issues in forensic and correctional research. *Aggression and Violent Behavior, 15*(6), 399-409. https://doi.org/10.1016/j.avb.2010.07.002

Ward, T., Melser, J., & Yates, P. M. (2007). Reconstructing the risk-need-responsivity model: A theoretical elaboration and evaluation. *Aggression and Violent Behavior, 12*(2), 208-228. https://doi.org/10.1016/j.avb.2006.07.001

Waskow, I. E. (1975). Selection of a core battery. *Psychotherapy Change Measures*, 245-269.

Webster, N., & Gardner, S. (2021). Offender personality disorder pathway Intensive Intervention Risk Management Service (IIRMS): Barriers to engagement and a vision for the future. *Probation Journal, 68*(1), 47-63. https://doi.org/10.1177/0264550520979363

West, L. (2014). Personality disorder & serious further offending. (Doctoral dissertation, University of Nottingham). https://core.ac.uk/download/33567380.pdf

Whewell, P., & Bonanno, D. (2000). The Care Programme Approach and risk assessment of borderline personality disorder. *Psychiatric Bulletin, 24*, 381-384. https://doi.org/10.1192/pb.24.10.381

Williams, M. M., Rogers, R., Sharf, A. J., & Ross, C. A. (2019). Faking good: An investigation of social desirability and defensiveness in an inpatient sample with personality disorder traits. *Journal of Personality Assessment, 101*(3), 253-263, https://doi.org/10.1080/00223891.2018.1455691

Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. G. B., Hadden, J., & Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS): Research and development. *The British Journal of Psychiatry*, *172*(1), 11-18. https://doi.org/10.1192/bjp.172.1.11

Wo, J. C., & Park, J. (2019). Places Not Cases? Rethinking Probation: The Effects of Nonprofit Organizations on Rearrest Rates of Probationers. *Crime & Delinquency*, *65*(5), 657-680. https://doi.org/10.1177/0011128718788046

World Health Organisation (WHO) (2018). ICD-11 International classification of diseases for mortality and morbidity statistics (11th revision). https://icd.who.int/browse11/l-m/en

World Prison Brief (2020, February, 14th). Institute for Crime & Justice Policy Research. https://www.prisonstudies.org/country/united-states-america.

Zhang, S. X., Roberts, R. E. L., & Callanan, V. J. (2006). Preventing parolees from returning to prison through community-based reintegration. *Crime & Delinquency*, *52*, 551-571. https://doi.org/10.1177/0011128705282594

Zhang, Z., Kim, H. J., Lonjon, G., & Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Annals of Translational Medicine, 7*(1), 16. https://doi.org/10.21037/atm.2018.12.10

Appendices

Appendix A

Scoping Review Search Terms

Table A1

Cochrane Library Initial Scoping Review Search Terms

Date	Search Terms	Limits	Hits
09.01.23	Title/Abstract/Keyword Search: ((Crim*) OR (offend*)) **AND** ((release*) OR (parole*) OR (probation*)) **AND** ((manag*) OR (supervis*) OR (rehabilita*) OR (treat*) OR (interven*) OR (diver*)) **AND** (("what works") OR (effectiveness) OR (efficacy) OR (outcome) OR (evaluat*))	Cochrane Reviews only	10

Date Limits: January 1990 – January 2023

Table A2

Cochrane Library Additional Scoping Review Search Terms

Date	Search Terms	Limits	Hits
09.01.23	Title/Abstract/Keyword Search: (reentry) OR (re-entry)	Cochrane Reviews only	4

Date Limits: January 1990 – January 2023

Table A3

Campbell Library Initial Scoping Review Search Terms

Date	Search Terms	Limits	Hits
09.01.23	(violen*) OR (offender*) **AND** ("post*release*") OR (release*) OR ("post*parole*") OR (parole*) OR ("after custod*") OR ("after prison") OR ("following custod*") OR ("following prison") **AND** ("offend*manag*") OR (supervis*) OR (rehabilita*) OR (treat*) OR (intervention) OR (diversion*) OR (probation) OR (therap*) **AND** (effective*) OR (efficacy) OR (outcome) OR ("what works")	Campbell Reviews only	2

Date Limits: January 1990 – January 2023

Table A4

Campbell Library Additional Scoping Review Search Terms

Date	Search Terms	Limits	Hits
09.01.23	Keyword Search: (reentry) OR (re-entry)	Campbell Reviews only	0

Date Limits: January 1990 – January 2023

Appendix B

Systematic Literature Review Search Terms by Database

Table B1

Search Terms Used for Systematic Literature Review by Database

Database	Date	Search Terms	Limits	Hits
PsycINFO	24.10.19	Subject Headings: exp. Violence OR exp Sex Offenses Keyword Search: ((Crim*) OR (offend*)) **AND** ((release*) OR (parole*) OR (probation*)) **AND** ((manag*) OR (supervis*) OR (rehabilita*) OR (treat*) OR (interven*) OR (diver*)) **AND** (("what works") OR (effectiveness) OR (efficacy) OR (outcome) OR (evaluat*))	(adulthood <18+ years> and "300 adulthood <age 18 yrs and older>" and ("0100 journal" or "0110 peer-reviewed journal" or "0120 non-peer-reviewed journal" or "0130 peer-reviewed status unknown") and english and male)	116
EMBASE	24.10.19	Subject Heading: exp violence/ OR exp sexual crime Keyword Search: ((Crim*) OR (offend*)) AND ((release*) OR (parole*) OR (probation*)) AND ((manag*) OR (supervis*) OR (rehabilita*) OR (treat*) OR (interven*) OR (diver*)) AND (("what works") OR (effectiveness) OR (efficacy) OR (outcome) OR (evaluat*))	(english and yr="1990 - Current" and (adult <18 to 64 years> or aged <65+ years>) and "humans only (removes records about animals)")	96
MEDLINE	24.10.19	Subject Heading: exp. violence/ OR exp sexual offenses Keyword Search: ((Crim*) OR (offend*)) AND ((release*) OR (parole*) OR (probation*)) AND ((manag*) OR (supervis*) OR (rehabilita*) OR (treat*) OR (interven*) OR (diver*)) AND (("what works") OR (effectiveness) OR (efficacy) OR (outcome) OR (evaluat*))	(("adult (19 to 44 years)" or "middle aged (45 plus years)") and english and humans and "humans only (removes records about animals)")	78
Web of Science	24.10.19	Topic Search: ((violen*) OR ("sexual offen*")) **AND** ((Crim*) OR (offend*)) **AND** ((release*) OR (parole*) OR (probation*)) **AND** ((manag*)	RESEARCH AREAS: (CRIMINOLOGY PENOLOGY OR PSYCHOLOGY OR BEHAVIORAL	317

		OR (supervis*) OR (rehabilita*) OR (treat*) OR (interven*) OR (diver*)) **AND** (("what works") OR (effectiveness) OR (efficacy) OR (outcome) OR (evaluat*))	SCIENCES OR PSYCHIATRY OR SOCIOLOGY OR REHABILITATION) AND Databases: (WOS) AND DOCUMENT TYPES: (ARTICLE) AND DOCUMENT TYPES: (ARTICLE) AND LANGUAGES: (ENGLISH)	
SCOPUS	24.10.19	Title/Abstract/Keywords: ((violen*) OR ("sexual offen*")) **AND** ((Crim*) OR (offend*)) **AND** ((release*) OR (parole*) OR (probation*)) **AND** ((manag*) OR (supervis*) OR (rehabilita*) OR (treat*) OR (interven*) OR (diver*)) **AND** (("what works") OR (effectiveness) OR (efficacy) OR (outcome) OR (evaluat*))	1990-current, (psychology, social science, arts and humanities), articles only, English language	230
NCJRS	24.10.19	Anywhere except Full Text Search: ((violen*) OR ("sexual offen*")) **AND** ((Crim*) OR (offend*)) **AND** ((release*) OR (parole*) OR (probation*)) **AND** ((manag*) OR (supervis*) OR (rehabilita*) OR (treat*) OR (interven*) OR (diver*)) **AND** (("what works") OR (effectiveness) OR (efficacy) OR (outcome) OR (evaluat*))	Narrowed by: decade: 1990 - 1999; 2000 - 2009; 2010 - 2019; Language: English; **Exclude all**: **Document type**: General Information; Government & Official Document; Statistics/Data Report; Book Chapter; Report; Case Study; Book; Instructional Material/Guideline; Reference Document; Technical Report; Interview; News; **Source type**: Books; Conference Papers & Proceedings; Blogs, Podcasts, & Websites; **Subject**: nij grant-related documents (11343); bja grant-related documents	355

			(11344); grants or contracts (03160); juvenile recidivism (09314); juvenile sex offenders (09320); violent juvenile offenders (08670); juveniles (03990); juvenile offenders (09505); female inmates (05072); female offenders (00411); juvenile gangs (04585)	
ASSIA	24.10.19	Anywhere except Full Text Search: ((violen*) OR ("sexual offen*")) **AND** ((Crim*) OR (offend*)) **AND** ((release*) OR (parole*) OR (probation*)) **AND** ((manag*) OR (supervis*) OR (rehabilita*) OR (treat*) OR (interven*) OR (diver*)) **AND** (("what works") OR (effectiveness) OR (efficacy) OR (outcome) OR (evaluat*))	NOT (subt.exact("juvenile offenders") AND yr(1990-2019))	157
Social Science Database	24.10.19	Anywhere except Full Text Search: ((violen*) OR ("sexual offen*")) **AND** ((Crim*) OR (offend*)) **AND** ((release*) OR (parole*) OR (probation*)) **AND** ((manag*) OR (supervis*) OR (rehabilita*) OR (treat*) OR (interven*) OR (diver*)) **AND** (("what works") OR (effectiveness) OR (efficacy) OR (outcome) OR (evaluat*))	1990-2019, (Article OR Review), NOT (juvenile offenders AND juvenile justice AND youth AND female AND women AND adolescent), English.	57

Date Limits: January1990 – January 2023

Appendix C

Additional Articles for Inclusion

The following articles were added to the literature database having been referenced in the papers returned by the initial search strategy:

- Adekeye & Emmanuel (2018)
- Clark (2015)
- Polaschek & Yesberg (2018)
- Ryan et al. (2022)
- Seewald et al. (2018)
- Travis & Petersilia (2001)
- Travis & Visher (2005)
- Wo & Park (2018)

Appendix D

List of Professionals Contacted and Example Contact Email

The following professionals were contacted directly to ensure a comprehensive examination of the literature was completed. Their responses are recorded:

- Dr Jackie Craissati – Consultant Clinical and Forensic Psychologist to the OPD Pathway – contacted by email. Response received.
- Professor Devon Polaschek – Clinical Psychologist and Joint Director of the New Zealand Institute for Security and Crime Science, University of Waikato – contacted by email. Response received.
- Dr Carine Lewis – OPD Pathway Research and Data Lead – contacted by phone. Response received.
- Professor Jason Davies – Consultant Clinical and Forensic Psychologist to the OPD Pathway – contacted by phone. Response received.
- Dr Julia Yesberg – Research Fellow in the Jill Dando Institute of Security and Crime Science – contacted by email. No response received.

Example email sent to professionals:

Dear _____,

I am currently undertaking a Doctorate in Applied Forensic Psychology at the University of Birmingham, England, and work as a Senior Forensic Psychologist in the Wales Offender Personality Disorder Pathway team. I am conducting a systematic literature review examining all relevant literature from 1990 to the present day that focuses on the process of re-entry into the community for high-risk men having served a prison sentence for violence or sexual violence. My research aims to examine the efficacy of Intensive Intervention and Risk Management Services (IIRMS) in the UK and the systematic literature review seeks to examine what is currently known about effective ways of managing, supporting and treating this cohort on re-entry into the community.

As a result of my searches and my knowledge of this area, I am getting in touch with practitioners and researchers who have been involved in this field of work to see if there is any additional literature that it would be useful to include in my review. If you are aware of any ongoing or unpublished literature which may not feature in typical academic databases, I would be very grateful if you could let me know. This will allow my search to be as comprehensive as possible and the review to be accurate and helpful.

If you would like any further details of either the systematic literature review or the research project, please don't hesitate to get in touch. I appreciate how busy you will be and am extremely grateful for any response you may be able to provide.

Kind regards,

Darius Cavina

Principal Forensic Psychologist

SE Wales Personality Disorder Pathway

Appendix E

Combined Results of CASP Quality Assessment Checklists

Table E1

Cohort Design CASP Checklist Results

Cohort Studies	1. Did the study address a clearly focused issue?	2. Was the cohort recruited in an acceptable way?	3. Was the exposure accurately measured to minimise bias?	4. Was the outcome accurately measured to minimise bias?	5. (a) Have the authors identified all important confounding factors?	5. (b) Have they taken account of the confounding factors in the design and/or analysis?	6. (a) Was the follow up of subjects complete enough?	6. (b) Was the follow up of subjects long enough?	7. What are the results of this study?	8. How precise are the results?	9. Do you believe the results?	10. Can the results be applied to the local population?	11. Do the results of this study fit with other available evidence?	12. What are the implications of this study for practice?	Score (/12)	%	Rating
Antonio & Crossett (2017)	Y	Y	Y	Y	Y	Y	Y	Y	N/A	N/A	Y	Y	Y	Y	12	100.0	Strong
Bouffard & Bergeron (2006)	Y	Y	Y	Y	Y	N	Y	Y	N/A	N/A	C/T	Y	Y	Y	10	83.3	Strong
Braga, Piehl & Hureau (2009)	Y	Y	C/T	Y	Y	Y	Y	Y	N/A	N/A	Y	Y	Y	Y	11	91.7	Strong
Hanson, Broom & Stephenson (2004)	Y	Y	C/T	Y	Y	N	Y	Y	N/A	N/A	C/T	Y	Y	Y	9	75.0	Moderate
Hedderman & Sugg (1996)	Y	Y	C/T	Y	C/T	Y	Y	Y	N/A	N/A	C/T	Y	Y	Y	9	75.0	Moderate
Lattimore & Visher (2009)	Y	Y	C/T	Y	Y	N	Y	Y	N/A	N/A	Y	C/T	Y	Y	9	75.0	Moderate
Procter (1996)	Y	Y	Y	Y	Y	Y	N	Y	N/A	N/A	Y	N	Y	Y	10	83.3	Strong

Y=Yes; N=No; C/T=Cannot Tell

Table E2

Randomised Control Trial Design CASP Checklist Results

RCT Studies	1.Did the trial address a clearly focused issue?	2.Was the assignment of patients to treatments randomised?	3.Were all of the patients who entered the trial properly accounted for at its conclusion?	4. Were patients, health workers and study personnel 'blind' to treatment?	5. Were the groups similar at the start of the trial?	6. Aside from the experimental intervention, were the groups treated equally?	7. How large was the treatment effect?	8. How precise was the estimate of the treatment effect?	9. Can the results be applied to the local population, or in your context?	10. Were all clinically important outcomes considered?	11. Are the benefits worth the harms and costs?	Score (/9)	%	Rating
Clark (2015)	Y	Y	Y	N	Y	Y	N/A	N/A	Y	Y	Y	8	88.9	Strong
Nathan, Centifanti, Baker & Hill (2019)	Y	Y	C/T	N	Y	Y	N/A	N/A	Y	Y	Y	7	77.8	Moderate
Shaul, Koeter & Schippers (2016)	Y	Y	Y	N	Y	Y	N/A	N/A	C/T	Y	Y	7	77.8	Moderate

Y=Yes; N=No; C/T=Can't Tell

Appendix F

Example Cohort Design CASP Form

CASP Checklist: 12 questions to help you make sense of a Cohort Study

How to use this appraisal tool: Three broad issues need to be considered when appraising a cohort study:

Are the results of the study valid? (Section A)
What are the results? (Section B)
Will the results help locally? (Section C)

The 12 questions on the following pages are designed to help you think about these issues systematically. The first two questions are screening questions and can be answered quickly. If the answer to both is “yes”, it is worth proceeding with the remaining questions. There is some degree of overlap between the questions, you are asked to record a “yes”, “no” or “can’t tell” to most of the questions. A number of italicised prompts are given after each question. These are designed to remind you why the question is important. Record your reasons for your answers in the spaces provided.

About: These checklists were designed to be used as educational pedagogic tools, as part of a workshop setting, therefore we do not suggest a scoring system. The core CASP checklists (randomised controlled trial & systematic review) were based on JAMA 'Users’ guides to the medical literature 1994 (adapted from Guyatt GH, Sackett DL, and Cook DJ), and piloted with health care practitioners.

For each new checklist, a group of experts were assembled to develop and pilot the checklist and the workshop format with which it would be used. Over the years overall adjustments have been made to the format, but a recent survey of checklist users reiterated that the basic format continues to be useful and appropriate.

Referencing: we recommend using the Harvard style citation, i.e.: *Critical Appraisal Skills*

Programme (2018). CASP (insert name of checklist i.e. Cohort Study) Checklist. [online]

Available at: URL. Accessed: Date Accessed.

Paper for appraisal and reference: Antonio & Crossett (2017)

Section A: Are the results of the study valid?

1. Did the study address a clearly focused issue?

Yes

~~Can't Tell~~

~~No~~

HINT: A question can be 'focused' in terms of

• the population studied

• the risk factors studied

• is it clear whether the study tried to detect a beneficial or harmful effect

• the outcomes considered

Comments: The research aims were clearly outlined and the population being studied and the control group were described in detail. It was clear that the aim of the research was to measure the effect of a cognitive behavioural intervention on recidivism using an untreated control group for comparison. Whilst the aim was to detect a beneficial effect of treatment, any harmful effects would also have been measured by the study design. The outcome measure of recidivism was clearly described.

2. Was the cohort recruited in an acceptable way?

Yes

~~Can't Tell~~

~~No~~

HINT: Look for selection bias which might compromise the generalisability of the findings:

• was the cohort representative of a defined population

• was there something special about the cohort

• was everybody included who should have been

Comments: Data was obtained from official records of those released from custody to parole supervision between Jan 2010 and May 2015. This set of data included over 87,000 convicted offenders and was used to create the intervention and the control groups. Most of these offenders were not referred for treatment but a matched control group was identified to ensure groups were as closely aligned as possible. The way missing data was managed was clearly described. One concern was that intervention non-completers were removed from the intervention group. It is known that treatment drop-outs often have a higher risk of reoffending than programme completers raising concerns that this may bias the results in favour of a treatment effect. This is fully acknowledged in the limitations section. Given the large numbers involved, the authors appear to have made significant steps to minimise selection bias.

Is it worth continuing?

3. Was the exposure accurately measured to minimise bias?

Yes

~~Can't Tell~~

~~No~~

HINT: Look for measurement or classification bias:

• did they use subjective or objective measurements

• do the measurements truly reflect what you want them to (have they been validated)

• were all the subjects classified into exposure groups using the same procedure

Comments: The intervention is clearly outlined in the paper and is described as a structured manualised programme delivering 64 hours of content with clear aims around improving thinking styles, anti-social attitudes, problem solving deficits, use of substances and employment opportunities. The referral process for participants entering the programme was clearly described.

4. Was the outcome accurately measured to minimise bias?

Yes

~~Can't Tell~~

~~No~~

HINT: Look for measurement or classification bias:

• did they use subjective or objective measurements

• do the measurements truly reflect what you want them to (have they been validated)

• has a reliable system been established for detecting all the cases (for measuring disease occurrence)

• were the measurement methods similar in the different groups

• were the subjects and/or the outcome assessor blinded to exposure (does this matter)

Comments: The primary outcome measure was identified as re-incarceration resulting from either the commission of a new crime and/or the violation of a technical condition of parole. The authors recognise in the limitations section that arrest rates are often a better proxy for reoffending but state that this data was not available. Outcome measurement was identical for both intervention and control groups, although it may have been the case that those who were referred for treatment were monitored more closely due to a belief that they were higher risk or because they had more conditions of parole due to treatment.

5. (a) Have the authors identified all important confounding factors?

Yes

~~Can't Tell~~

~~No~~

HINT:

• list the ones you think might be important, and ones the author missed

Comments: The authors recognise the confounding variables that can exist when two different groups of participants are compared. For example, it is entirely possible that those referred for treatment were considered higher risk and therefore may have had worse outcomes anyway, regardless of treatment. Another potential confounding variable was participants prior programming experience. It is possible that one or other of the groups had significantly more treatment in the past, making it

harder to attribute any results solely to the intervention under evaluation. Authors also recognise that by removing programme non-completers, they may have introduced a degree of selection bias.

5. (b) Have they taken account of the confounding factors in the design and/or analysis?

Yes

~~Can't Tell~~

~~No~~

HINT:

• look for restriction in design, and techniques e.g. modelling, stratified-, regression-, or sensitivity analysis to correct, control or adjust for confounding factors

Comments: The authors recognised the risk of confounding variables when using different groups for comparison. To minimise this risk, they match the intervention and the control groups as closely as possible using gender, age, race, location of release, offence category and history of violence. They also used logistic regression to highlight those variables which were predictive of being referred to the intervention group to ensure that participants in the control group were equally likely to have been referred for treatment. Excluding programme non-completers and not controlling for previous treatment were not able to be accounted for in the design, but these issues were fully discussed in the limitation section.

6. (a) Was the follow up of subjects complete enough?

Yes

~~Can't Tell~~

~~No~~

HINT: Consider

• the good or bad effects should have had long enough to reveal themselves

• the persons that are lost to follow-up may have different outcomes than those available for assessment

• in an open or dynamic cohort, was there anything special about the outcome of the people leaving, or the exposure of the people entering the cohort

6. (b) Was the follow up of subjects long enough?

Yes

~~Can't Tell~~

~~No~~

Comments: As follow up was measured by re-incarceration rates, there were no difficulties in gathering relevant data from all participants. The follow up period was 36 months which is generally considered a reasonable period to demonstrate either positive or negative effects of the intervention. In terms of the potential selection bias caused by removing non-completers from the intervention group, there is some concern that this sub-group may have had different outcomes to the rest of the intervention group. This is fully discussed in the limitations section.

Section B: What are the results?

7. What are the results of this study?

HINT: Consider

• what are the bottom line results

• have they reported the rate or the proportion between the exposed/unexposed, the ratio/rate difference

• how strong is the association between exposure and outcome (RR)

• what is the absolute risk reduction (ARR)

Comments: Results indicate that, holding all other covariates constant, the intervention group reduced their risk of recidivism by 24% compared with the control group. When examining high-risk offenders, the results were even more favourable with the intervention group reducing their risk of recidivism by 31% compared with the control group. However, the intervention group were returned to custody for a new offence at a greater rate than the control group. The results suggest that there is a strong association between intervention and outcome but questions the utility of the outcome measure as it includes both technical breaches and new offending.

8. How precise are the results?

HINT:

• look for the range of the confidence intervals, if given

Comments: The results are presented clearly and precisely with 95% confidence intervals provided. The range of the confidence intervals is relatively large but they still equate to a minimum 15% reduction in recidivism and a maximum 33% reduction in recidivism for the entire group.

9. Do you believe the results?

Yes

~~Can't Tell~~

~~No~~

HINT: Consider

• big effect is hard to ignore

• can it be due to bias, chance or confounding

• are the design and methods of this study sufficiently flawed to make the results unreliable

• Bradford Hills criteria (e.g. time sequence, dose-response gradient, biological plausibility, consistency)

Comments: The effect is relatively large and the results are presented clearly. However, the two factors which make up the outcome measure - new criminal offence or technical breach - makes the effect difficult to attribute to the programme alone. This doesn't mean that the results are invalid, simply that they require interpretation.

Section C: Will the results help locally?

10. Can the results be applied to the local population?

Yes

~~Can't Tell~~

~~No~~

HINT: Consider whether

• a cohort study was the appropriate method to answer this question

• the subjects covered in this study could be sufficiently different from your

population to cause concern

• your local setting is likely to differ much from that of the study

• you can quantify the local benefits and harms

Comments: A cohort design was an appropriate method for this study and the large group sizes mean that they are likely to be representative of other groups of medium and high-risk released offenders.

11. Do the results of this study fit with other available evidence?

Yes

~~Can't Tell~~

~~No~~

Comments: To some degree the results fit with other findings in this area. However, this is a difficult cohort to study making results from different studies unpredictable.

12. What are the implications of this study for practice?

Yes

~~Can't Tell~~

~~No~~

HINT: Consider

• one observational study rarely provides sufficiently robust evidence to recommend changes to clinical practice or within health policy decision making

• for certain questions, observational studies provide the only evidence

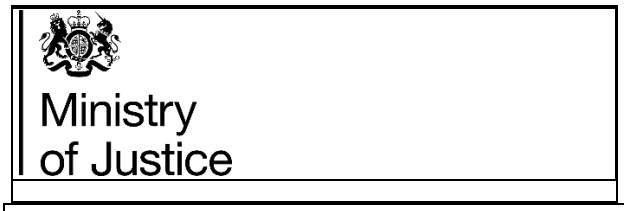
• recommendations from observational studies are always stronger when supported by other evidence

Comments: As the sample size is relatively large and the design is robust and accounts for some of the confounding variables, this paper is a very helpful contribution to the literature in this area. However, it is important that the results are clearly understood in terms of the treatment effect appearing to be limited to technical breaches rather than reoffending. The results of this study will help shape future research in this area when combined with other observational papers.

Appendix G

Memorandum of Understanding

The following Memorandum of Understanding is between the NHS / HMPPS OPDP National Research and Data Team and NHS / HMPPS OPDP Team in Wales



Memorandum of understanding between the NHS / HMPPS OPDP National Research and Data Team and NHS / HMPPS OPDP Team in Wales.

Management of shared data

DATED the 10th day of May 2021

BETWEEN

Offender Personality Disorder Pathway Research & Data Team, Rehabilitation & Support Services Group, Directorate of Safety and Rehabilitation, Ministry of Justice, 1st Floor, Southern House, Croydon, CR0 1XG.

AND

Darius Cavina, Principal Forensic Psychologist for the Offender Personality Disorder Pathway in Wales (NHS and HMPPS co-funded project), 33 Westgate Street, Cardiff, CF10 1JE.

Collectively hereafter "the Parties"

1 Purpose

1.1 The purpose of this Memorandum of Understanding is to establish the standards and processes to protect information that the parties intend to share in order to enable Darius Cavina from the joint NHS and HMPPS Offender Personality Disorder Pathway (OPDP) team to

complete a research project looking at the effective management of high-risk violent offenders in the community. In addition to supporting the business priorities of HMPPS and the OPDP team in reducing reoffending and increasing public safety, this project forms part of a doctorate programme of study being completed by Darius Cavina at the University of Birmingham.

2 Description

2.1 This MoU will assist Darius Cavina, of the joint NHS / HMPPS OPDP team, to complete a study of secondary data, examining the relationship between community management strategies, such as Intensive Intervention and Risk Management Services (IIRMS), personality disorder traits and recall to custody among high-risk violent offenders. The proposed data share is to allow analysis of the relationship between the choice of community management strategy, personality disorder traits and recall. This will involve matching recall, breach and warning data to an original dataset of men who have completed one of three community management processes. The broad aims of the study are to identify whether any of the management strategies are particularly effective in keeping high-risk violent offenders safely in the community for an extended period and whether this is influenced by the type of offender being studied, with reference to age, offence type and personality disorder traits. The research links to HMPPS business priorities of reducing reoffending.

2.2 This MoU concerns Darius Cavina, of the NHS / HMPPS OPDP team, being provided with a Ministry of Justice (MoJ) nDelius data extract as needed above. The project requests a one-off data share between MoJ and the NHS / HMPPS OPDP team in Wales to allow the research to take place. The research has been approved by the MoJ National Research Committee (Ref: 2019-119), has received approval from the University of Birmingham's Ethical Review Committee (Ref: ERN_19-1244) and is fully supported by the OPDP Research and Data team. Darius Cavina is requesting nDelius data (recall, breach and warning data) for a cohort of male offenders who were managed through one of three community supervision processes (Intensive Intervention and Risk Management Services, Integrated Response Integrated Services and Probation as Usual) between April 2014 and March 2019. Previous research into this complex group of high-risk men is mixed, although there is some indication that more intensive supervision and support processes can lead to extended periods in the community.

2.3 The data share will not involve personal details of individuals, as the data set will have been anonymised by the central OPDP team by the time it is shared with Darius Cavina. Analysis of the data shared will provide aggregated 5-year recall, breach and warning information so as to identify those variables, both individual and service level, which are most associated with subsequent community transgressions or recalls to custody.

2.4 The data share will consist of the following steps of which the Parties are both in agreement and acknowledge their respective obligations.

2.5 The OPDP Research and Data Team will send anonymised data sets to Darius Cavina containing individual level information regarding men who have been managed by IIRMS, IRiS and probation as usual. Darius Cavina will then use Propensity Score methods to identify a suitable control group using the following variables:

- Offence type
- Sentence type
- Sentence length

- DoB
- Ethnicity
- Release from prison year
- OGRS 1 score
- OGRS 2 score
- OGP score
- OVP score
- ROSH score
- RSR score
- OPDP screening variables

2.6 Once the matching process has been completed, the anonymised data will be sent back to the OPDP Research and Data Team who will use the key that only they hold to reinsert identifiers in order to request recall, breach and warning data from MoJ. This final data set will be returned to the OPDP Research and Data Team who will remove any identifiable information and send the anonymised data set back to Darius Cavina for analysis. The initial dataset submitted to MoJ will only include those variables needed to identify the individual for matching purposes including:

- CRN Number
- NOMIS ID

2.7 nDelius variables being requested include the following offence-level and descriptive data variables:

- Offence ID
- Offence start date
- Court / Caution Date
- Disposal Type
- Disposal Category
- Disposal Duration
- Primary Offence Flag
- DoB
- Ethnicity
- OPD Pathway Screening information (in Personal Circumstances)
- MAPPA Level
- Release Date

- Service Start Date (IIRMS or IRiS)
- Service End Date (IIRMS or IRiS)
- Recall Type
- Recall Date
- Breach Type
- Breach Date
- Warning Type
- Warning Date

2.8 Only one researcher within the joint NHS / HMPPS OPDP team in Wales will have access to anonymised MoJ data. This researcher is Darius Cavina, Principal Forensic Psychologist employed by OPDP in Wales. The OPDP Research and Data Team will have access to unanonymised individual level data and will not be involved in the analysis phase. Both parties agree to comply with all information assurance protocols and other expectations of MoJ account holders. This project is supervised by Dr Caroline Oliver, Course Director for the Doctorate in Forensic Psychology Practice at the University of Birmingham, although she will not have access to individual level data. An agreement is in place that Dr Oliver will have access to aggregated anonymised data once the analysis is complete, in order to supervise the research. No raw or individual level data will be shared or transferred to Dr Oliver or anyone else not named in this MoU. Anonymised data shared between the OPDP Research and Data Team and Darius Cavina will be transferred via secure email ('.justice.gov.uk') and will remain on a password protected secure MoJ laptop/desktop only accessible by the primary researcher.

2.9 The overarching principles agreed to by both parties is that they will comply with the:

- Data Protection Act (2018)
- HMG Security Policy Framework
- The Information Commissioner's Framework Code of Practice for sharing personal information available at: https://ico.org.uk/media/2615361/data-sharing-code-for-public-consultation.pdf

2.10 This includes:

2.10.1 Preserving the integrity, confidentiality and availability of shared data and preventing the corruption or loss of such data.

2.10.2 Processing the data only to the extent, and in such manner, as is necessary for the effective functioning of the justice system or as is required by law or any regulatory body.

3 Specific measures to protect personal information

3.1 The primary researcher from the NHS / HMPPS OPDP team in Wales shall at all times aim to protect personal data whose release or loss could cause harm or distress to individuals from compromise of its confidentiality, integrity or availability.

3.2 Following the matching of data to nDelius records, the primary researcher will only have access to anonymised data, which will be entered into SPSS for analysis. Within SPSS each individual participant has a unique identifier, in place of any personal identifying information. The final report will only contain aggregated data and it will not be possible to identify any individuals in the sample. The final research report will be shared amongst HMPPS and NHS stakeholders for information, with the aim of supporting the business priorities of reducing reoffending and increasing public safety. No individual level data will be shared with anyone not named in this MoU.

3.3 The legal basis for sharing this information is contained within section 14 of the Offender Management Act 2007, and the DPA 2018. Any sharing of personal data will be done in accordance with the provisions of the Data Protection Act 2018 (DPA). The DPA provides a framework for how personal data should be correctly handled. The DPA neither promotes nor prohibits the sharing of personal data, but its principles apply to data sharing as they apply to any other form of processing of personal data.

3.4 The shared information is proportionate to the purpose of the project.

3.5 All nDelius data will be viewed, stored and processed within the Dom1 server used by both MoJ and OPDP. NHS / HMPPS OPDP teams are hosted by the Ministry of Justice and staff members within OPDP teams are Ministry of Justice civil servants bound by departmental security protocols and procedures, the Official Secrets Acts and the Civil Service Code. The data will be held securely on the Dom1 server in a folder that only the researchers named in this MoU will have access to. The folder will be password protected.

3.6 The NHS / HMPPS OPDP team shall handle all information that meets the definition as set out in HMG Security Classification system definitions for OFFICIAL according to the prescribed standards including the specific measures set out in this MoU.

3.7 The NHS / HMPPS OPDP team agrees to entirely avoid the use of removable media for storage for access to personal data. All personal data will be stored on the Dom1 server and will be transferred when required using MoJ ICT systems, such as '.justice.gov.uk' mail.

3.8 All material that has been used for shared protected data will be subject to controlled disposal (if it is agreed that the material is not/cannot be returned to the source party).

3.8.1 Paper records will be destroyed by incineration, pulping or shredding so that reconstruction is unlikely.

3.8.2 Electronic media will be disposed of through secure destruction, overwriting, erasure or degaussing for re-use.

3.9 Shared data will be disposed of 5 years post report completion and assessment. The timescales for completion of analysis are 6 months from when access to the final anonymised data set is granted and submission for publication within a further period of 12 months. Estimated timescale for destruction of data is March 2027. Outcomes of statistical analyses may be appended to the report, but these will not include identifiable information.

3.10 The primary researcher from the NHS / HMPPS OPDP team is aware of the sensitivity of the data, the handling requirements, and has successfully completed information risk awareness training recently and is required to do so annually.

3.11 The primary researcher from the NHS / HMPPS OPDP team shall maintain at all times the 'need-to know' principle and limit access to information records to the minimum possible.

3.12 The primary researcher from the NHS / HMPPS OPDP team shall maintain business continuity and disaster recovery frameworks that cover the protection of shared data to maintain its integrity, availability and confidentiality.

3.13 The primary researcher from the NHS / HMPPS OPDP team shall not process personal data provided by another party outside the European Economic Area.

3.14 The primary researcher from the NHS / HMPPS OPDP team is aware of and entirely supportive of the principles of data protection in law and policy. This data share complies with the data protection principles of the Data Protection Act 2018.

4. Requirement

4.3 The primary researcher from the NHS / HMPPS OPDP team requires access to nDelius data to identify recall, breach and warning data for the project specified above.

4.4 The data transfer is described in section 2.5.

5. Reporting

5.1 If at any time the NHS / HMPPS OPDP team suspect or have reason to believe that shared data has or may become corrupted, lost or sufficiently degraded in any way for any reason, then they shall notify MoJ Analytical Services, MoJ Operational Security and MoJ Corporate Security Branch immediately.

5.2 Where equivalent measures to the specified requirements are adopted, or in exceptional

circumstances in which measures cannot be applied to systems that will be used to hold the shared data, the NHS / HMPPS OPDP team will advise the other party through the nominated contact of those exceptions or equivalent measures that have been put in effect.

5.3 Annually, the NHS / HMPPS OPDP team shall monitor disposal channels for paper or electronic media being disposed of.

5.4 The NHS / HMPPS OPDP team acknowledges that it is each party's obligation to maintain an adequate and effective system of controls to govern data within their organisations and that the MoJ place reliance on those controls as the underpinning of this MoU. Significant issues identified either through the controls set out in this section or through other processes of internal review should be advised to the other party where the weakness relates to the handling of data provided by the other party to the MoU.

6. Nominated contacts

6.1 Reports produced under section 3 above shall be provided to the Parties' nominated Information Risk Owners as set out in the table below.

	Name	Contact details

NHS / HMPPS Offender Personality Disorder Pathway Team	Name: Aisling O'Meara Title: Evidence Specialist	OPDP Research and Data Team, Rehabilitation & Support Services Group, Directorate of Safety and Rehabilitation, 1st Floor, Southern House, Croydon, CR0 1XG. m: e:
NHS / HMPPS Offender Personality Disorder Pathway Team in Wales	Name: Darius Cavina Title: Principal Forensic Psychologist	Offender Personality Disorder Pathway in Wales, 33 Westgate Street, Cardiff, CF10 1JE m: e:

SIGNED for and on behalf of

Ministry of Justice

By: Aisling O'Meara

SIGNED for and on behalf of

NHS / HMPPS OPDP Team in Wales

By: Darius Cavina

Annex

Risk Management Plan

Risk Management

There will be a memorandum of understanding in place between OPDP Research and Data Team (part of MoJ) and Darius Cavina of the NHS and HMPPS OPDP team in Wales in order to manage the risks associated with this data share.

Possible risks identified were:

- Inadequate disclosure controls increase the likelihood of information being shared inappropriately.
- The context in which information is used or disclosed can change over time, leading to it being used for different purposes without people's knowledge.
- Data loss
- Failure to implement appropriate technological security measures
- Failure to ensure that staff are properly trained and are aware of potential privacy risks.
- Failure to develop ways to safely anonymise the information when it is possible to do so.
- Failure to produce data share agreements which make clear what information will be shared, how it will be shared and who it will be shared with.

Risk Mitigation

In order to mitigate against the risks as identified, a memorandum of understanding is in place between the HMPPS OPDP National team (part of MoJ) and the NHS and HMPPS OPDP team in Wales.

Appendix H

Variables Requested from North of England Commissioning Support Unit (NECSU)

Table H1

Variable No	Variable Name	Variable Category	Variable Purpose
1	Case Identifier	Indentifier	Indentifier
2	Service Name	Group	Group Allocation
3	Gender	Demographic Data	Matching Variables
4	Date of Birth		
5	Ethnicity Group		
6	Main Offence (Offence Description)	Offence Data	
7	Offence Category		
8	Sentence Type		
9	Sentence Length (months)		
10	First Release From Custody Date (after study start date)		
11	OGRS1 Score (at time of release)	Risk Data	
12	OGRS2 Score (at time of release)		
13	OGRS Assessment Date		
14	RoSH Category		
15	Oasys/DSPD Items	OPDP Screening Data	
16	Childhood Difficulties		
17	Mental Health		
18	Self-Harm/ Suicide Attempts		
19	Challenging Behaviour		
20	OPDP Complexity Score		
21	Count of Warnings	Warning Data	Outcome Variables
22	First Warning Type		
23	First Warning Date		
24	Latest Warning Type		
25	Latest Warning Date		
26	Count of Breaches	Breach Data	
27	First Breach Type		
28	First Breach Date		
29	Latest Breach Type		
30	Latest Breach Date		
31	Recall Durnig Follow Up Period? (Y/N)	Recall Data	
32	Reason for Recall		
33	Recall Date		
34	Reconviction During Follow Up Period? (Y/N)	Reconviction Data	
35	Reconviction Offence Type		
36	Reconviction Date		

Appendix I

Additional Explanation of the Propensity Score Method Used

With numerous propensity score methods available, it is essential that consideration is given to the most appropriate method for the proposed study. There follows a detailed description of the decisions made in the current study, with justifications offered at each stage.

Matching groups using the propensity score, or similar distance estimate, eliminates or reduces the bias when calculating the treatment effect (Austin, 2007), can account for numerous confounding variables (Kainz et al., 2017) and lends itself to criminology research, which predominantly adopts an observational approach (e.g., Balusek, 2014; Gibson et al., 2012; Kim & Clark, 2013). Identifying the control group without access to outcome data ensures objectivity of the study design, further reducing the risk of selection bias (Rubin, 2007). In the current study, matched control groups were identified for both intervention samples, as well as for a combined intervention sample, prior to examination of the outcome data.

Covariate selection is essential to the success of the matching process, as it determines the amount of bias that can be removed from the sample (Shadish, 2013). The aim in covariate selection is to satisfy the strongly ignorable treatment assignment (SITA) assumption (Rosenbaum & Rubin, 1983). That is, to ensure that all covariates that may have an impact on treatment assignment and outcome are included in the matching process (Shadish & Steiner, 2010). Those covariates that are unrelated to treatment assignment and outcome are considered ignorable (Kainz et al., 2017). Whilst there is no agreed consensus in the literature regarding appropriate covariates to use in the matching process (Austin, 2011) and no way of testing whether the SITA assumption is met (Shadish, 2013; Shadish & Steiner, 2010), it is essential that every effort is made to include all covariates that could have a potential impact on treatment selection and outcome. The literature guides researchers to rely on subject-matter expertise and a review of the existing literature to identify relevant covariates for matching (Shadish, 2013; Shadish & Steiner, 2010), whilst paying closest attention to those covariates that would cause concern in an RCT study if baseline imbalance was detected (Austin, 2014). Covariates should be assessed before any exposure to treatment, to ensure the treatment itself does not influence measurement of the covariate (Kainz et al., 2017). Further, it is recommended that substantial thought is given to the causal model that explains the outcome, in order that all relevant covariates are included (Pearl, 2009). Covariates must also be accurately measured, as poorly approximated measures may not only fail to reduce bias but may also introduce additional bias into the model (Pearl, 2009; Steiner et al., 2011).

Table I1 shows the covariates used in a sample of relevant literature, as well as the covariates used in the current study. The covariates used for matching in the current study were chosen following an extensive review of the literature and were guided by subject-matter expertise held by the primary researcher and other professionals within the Ministry of Justice. Basic demographic details and release from prison dates were included to ensure similar groups of individuals were compared at similar times, to account for political, funding or decision-making differences across time. To ensure the study compared men with equitable levels of risk, a range of offence and risk indicators were used as matching variables. The Offender Group Reconviction Scale version 3 (OGRSv3) is a predictor of reoffending based on static data related to age, gender and criminal history (Francis et al., 2007). OGRSv3 has been used

Table I1

Matching Criteria Used in Relevant Literature, Including the Current Study

Study	Method	Matching Covariates
Antonio & Crossett (2016)	Propensity score matching	Gender, age, race, location of release, offence category, history of violence and Level of Service Inventory-Revised (LSI-R) score (risk level).
Braga, Piehl & Hureau (2009)	Propensity score matching	Age, race, current conviction offence, past gang association/membership and past criminal arrest history.
Clark (2015)	Randomised control trial (2:1 allocation of treatment to control groups)	Random assignment to groups, although numerous cases were excluded from both groups *after* randomisation because they didn't meet eligibility criteria.
Jolliffe et al. (2017)	One-to-one propensity score matching	Age, time between assessments, initial risk of harm rating, OGP level, OVP level and sentence type.
Lattimore & Visher (2010)	Mixed randomised control trial (two adult sites) and weighted propensity score method (10 adult sites) design.	Weighted analysis based on housing, employment, family/peer/community involvement, substance use, physical and mental health, criminal behaviour and recidivism.
Nathan et al. (2019)	Stratified randomised control trial	Stratified by type of offence, diagnosis of drug/alcohol abuse and designated probation office
Pearce et al. (2017)	Stratified randomised control trial	Stratified by age, gender and baseline service utilisation
Procter (1996)	Case-by-case matched control group design	Age, previous / current convictions, age of victim, gender of victim, force used, genital to genital contact, length of time offender was followed up for.
Current Study	One-to-many propensity score matching	Offence type, sentence type, sentence length, age, ethnicity, risk indicators (OGRS 1 and 2 and RoSH), year of release, and OPDP complexity score.

within the Probation Service since 2008, has been validated for use with prisoners by Whiting (cited in Howard et al., 2009) and can predict group reconviction rates both one year (OGRS1) and two years (OGRS2) post-release (Howard et al., 2009).

In addition to a reoffending predictor, the current study is also interested in the potential severity of reoffending. The Risk of Serious Harm (RoSH; HMPPS, 2022) measure is a structured clinical judgement tool using both static and dynamic risk factors to assess both the imminence and severity of a potential further offence. The four categories of risk are *low*, *medium*, *high* and *very high* and a comprehensive guide is provided to professionals when making their judgements (HMPPS, 2022). In addition to predictors of both likelihood and severity of potential reoffending, the Offender Personality Disorder Pathway (OPDP) have developed a screening tool which identifies personality complexities likely to affect progress through the Criminal Justice System (CJS). These variables have been established as robust indicators of personality difficulties among violent and sexually violent offenders (Craissati et al., 2008; Minoudis et al., 2012). The variables used as proxy measures for complexity include anti-social traits, mental health difficulties, childhood trauma, self-harm / suicide attempts and persistent problematic behaviour. OPDP screening data is used in the current study to ensure matches between cases with equitable levels of personality complexity. The triangulation of risk and complexity data across three separate measures is considered a robust method of matching cases across intervention and control groups with similar likelihood, imminence, and severity of potential reoffending.

Prior to choosing a matching method for observational studies, it must first be decided for which population the treatment effect is being estimated. Choosing the correct *estimand* is essential, as this often dictates the type of matching method used (Greifer & Stuart, 2021). The current study is interested in the effect of treatment on those in the IIRMS, IRiS and combined Intervention groups. Put simply, what would the difference in outcome be if those who were offered treatment were, in fact, not offered treatment. This means that the current study uses the average treatment effect in the treated (ATT) as its estimand. This works particularly well in the current study as there are significantly more control cases than intervention cases to select from, thus increasing the likelihood of well-matched samples (Stuart, 2010). Guidance from the literature suggests that nearest-neighbour matching without a caliper (Stuart, 2010), full matching (Hansen & Klopfer, 2006), fine stratification (Desai et al., 2017) and genetic matching (Diamond & Sekhon, 2013) are appropriate methods of estimating the ATT. Fine stratification is not considered in the current study, as it is suggested by Desai et al. (2017) that this method should only be used if the incidence rate is <5%, which is not the case here.

When matching using propensity score techniques, a series of decisions are required about the specific statistical methods to use. Some of these decisions can be made in advance, depending on the design and aims of the study. For example, deciding whether intervention or control cases will be used multiple times can be done in advance. The current study uses each case once, to ensure that the outcome of a single individual does not unduly bias the results. This is made possible by the fact that the control group is significantly larger than the treatment groups, meaning that cases are not required to be used multiple times. Other decisions should be made once the quality of covariate balance has been assessed across numerous methods, but before the treatment effect is estimated (Greifer, 2022b).

In order to select the methodology that provides the best balance between groups, it is important to examine data indicative of balance both before and after matching. Balance is an assessment of the similarity of the distribution of all covariates across intervention and control groups (Stuart, 2010). There are a variety of methods of assessing balance and it is recommended that a combination of methods is used to ensure the accuracy of the matching method (Greifer, 2022a). Comparing standardised mean difference (SMD) scores for each covariate before and

after matching allows us to examine the improvement in balance as a result of the chosen methodology. SMDs approximating zero after matching indicate minimal difference across groups for individual covariates. Belitser et al. (2011) offer support for the use of SMD scores as measures of covariate balance and Ali et al. (2014) conclude it should be the measure of choice when assessing covariate balance. Comparing the variance of each covariate in one group to that in the other – the variance ratio – is another helpful means of assessing covariate balance after matching (Austin, 2009). Variance ratios approximating 1 after matching indicate similar variances across groups, with ratios below 2 being considered an acceptable indication of balance (Zhang et al., 2019). Empirical cumulative density function (eCDF) statistics, measured by the Kolmogorov-Smirnov statistic, allow for an assessment of balance across the entire distribution of a covariate, rather than just the mean or variance (Greifer, 2022a). This provides useful data in the assessment of balance, in addition to SMD and variance ratio statistics (Austin & Stuart, 2015). eCDF scores approximating zero after matching indicate good covariate balance.

In addition to statistical data, it can be helpful, and often more accessible, for visual representations to be displayed (Stuart, 2010). Love Plots displaying before and after matching Standardised Mean Difference (SMD) and Kolmogorov-Smirnov (K-S) statistics are the most helpful and accessible visual comparison tools, allowing for a threshold to be set for SMD and K-S scores. This facilitates a quick method of assessing which specific variables create most imbalance across groups, thus enabling the selection of a method of covariate balance that minimises these differences (Ho et al., 2007; Austin, 2009).

Final Matching Methods

Table I2 shows the final matching methods used and the sizes of the intervention and control groups across each sample. Figures I1-3 show Love Plots providing a visual representation of the covariate balance, assessed by absolute Standardized Mean Difference (SMD) and Kolmogorov-Smirnov (K-S) statistics, across the IIRMS, IRiS, and Intervention groups. As can be seen, the absolute SMD and K-S statistics for the matched samples are within acceptable tolerances of 0.1 for each covariate, indicating excellent covariate balance between groups (Greifer, 2022a). Specific balance data for every covariate within each group can be viewed in Tables J1-3 of Appendix J. These data were used to justify the selection of each final matching method.

Table I2

Final Matching Methods and Sample Sizes

	IIRMS	IRiS	Intervention
Matching Method	Nearest Neighbour Matching on the Propensity Score	Genetic Matching on the Propensity Score	Genetic Matching on the Propensity Score
Matching Ratio (intervention:control)	1:4	1:4	1:4
Intervention Cases (% of available sample)	373 (100)	75 (100)	448 (100)
Control Cases (% of available sample)	1,492 (39.7)	300 (8.0)	1,792 (47.7)

Figure I1

Love Plot of Covariate Balance for IIRMS Population

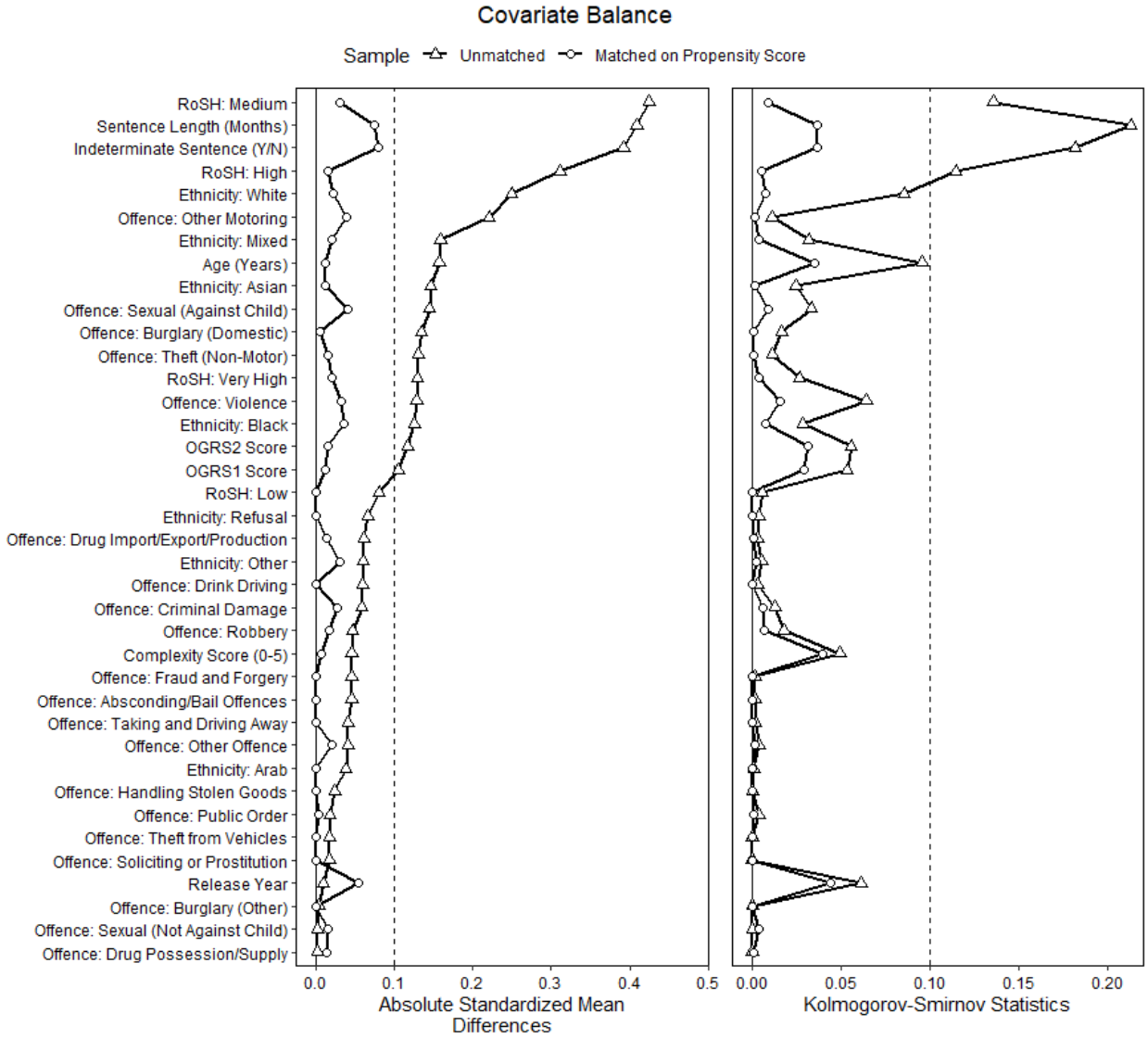


Figure I2

Love Plot of Covariate Balance for IRiS Population

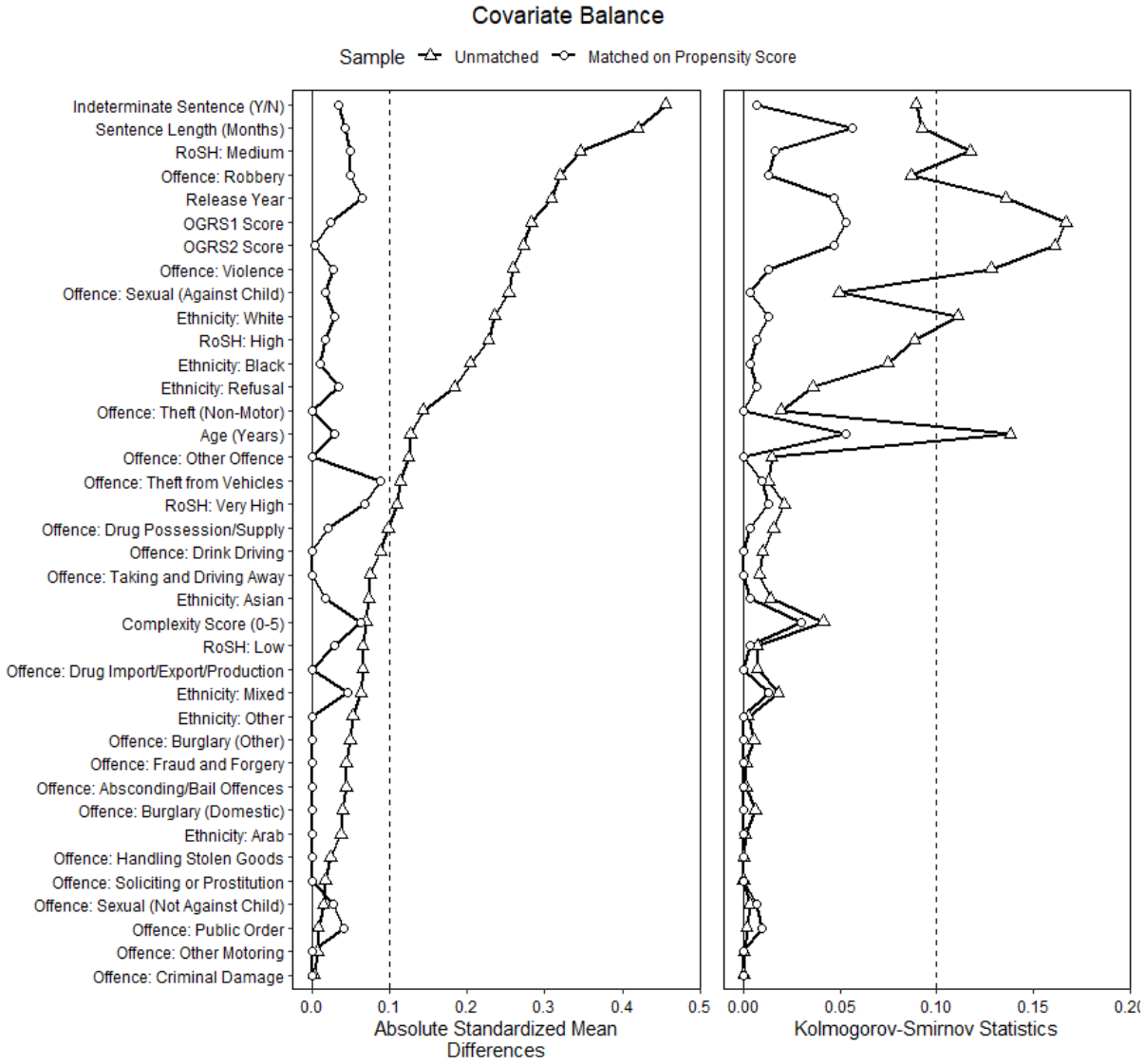
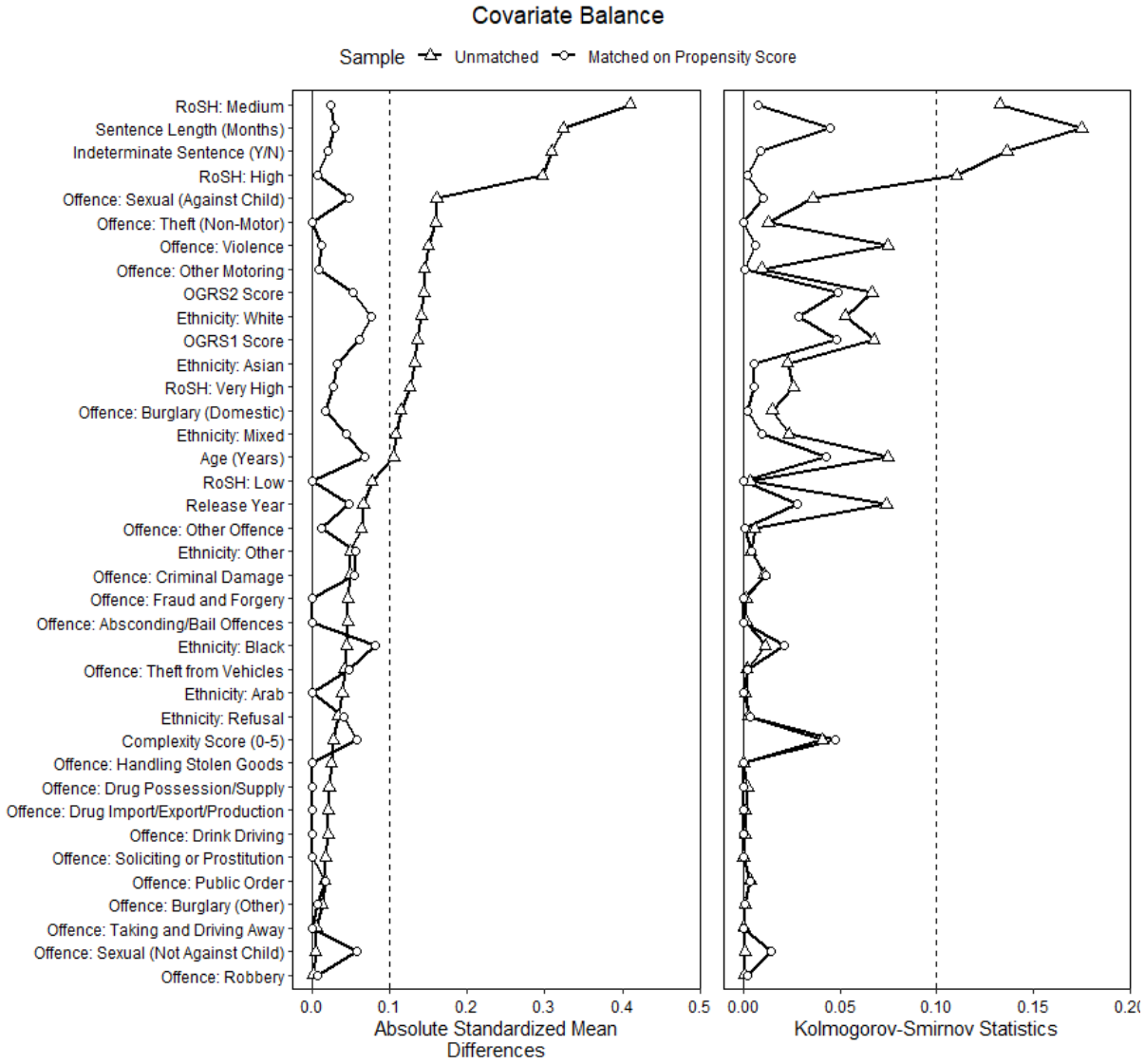


Figure I3

Love Plot of Covariate Balance for Intervention Population



Appendix J

Balance Statistics Before and After Matching

Table J1

Balance Statistics Before and After Matching for IIRMS Population

	Unmatched			Matched		
	Standardised Mean Difference	Variance Ratio	Kolmogorov-Smirnov Statistic	Standardised Mean Difference	Variance Ratio	Kolmogorov-Smirnov Statistic
Age (Years)	0.157	0.908	0.096	0.011	0.881	0.036
Release Year	-0.010	0.781	0.062	0.054	0.770	0.044
Ethnicity: Arab	-0.038	.	0.001	0.000	.	0.000
Ethnicity: Asian	-0.147	.	0.025	0.012	.	0.002
Ethnicity: Black	-0.125	.	0.029	-0.035	.	0.008
Ethnicity: Mixed	-0.159	.	0.032	-0.020	.	0.004
Ethnicity: Other	0.060	.	0.005	0.030	.	0.003
Ethnicity: Refusal	-0.066	.	0.004	0.000	.	0.000
Ethnicity: White	0.250	.	0.086	0.022	.	0.007
Offence: Absconding/Bail Offences	-0.045	.	0.002	0.000	.	0.000
Offence: Burglary (Domestic)	-0.134	.	0.017	0.005	.	0.001
Offence: Burglary (Other)	0.004	.	0.000	0.000	.	0.000
Offence: Criminal Damage	0.059	.	0.013	0.027	.	0.006
Offence: Drink Driving	-0.059	.	0.003	0.000	.	0.000
Offence: Drug Import/Export/Production	-0.061	.	0.003	0.013	.	0.001
Offence: Drug Possession/Supply	-0.002	.	0.000	-0.013	.	0.001
Offence: Fraud and Forgery	-0.045	.	0.002	0.000	.	0.000
Offence: Handling Stolen Goods	-0.024	.	0.001	0.000	.	0.000
Offence: Other Motoring	-0.221	.	0.011	0.039	.	0.002
Offence: Other Offence	-0.041	.	0.004	-0.020	.	0.002
Offence: Public Order	-0.018	.	0.004	-0.003	.	0.001
Offence: Robbery	0.047	.	0.018	-0.017	.	0.007
Offence: Sexual (Against Child)	-0.145	.	0.033	-0.041	.	0.009
Offence: Sexual (Not Against Child)	-0.002	.	0.001	-0.016	.	0.004
Offence: Soliciting or Prostitution	-0.017	.	0.000	0.000	.	0.000
Offence: Taking and Driving Away	-0.041	.	0.002	0.000	.	0.000
Offence: Theft (Non-Motor)	-0.130	.	0.012	-0.015	.	0.001
Offence: Theft from Vehicles	-0.017	.	0.000	0.000	.	0.000
Offence: Violence	0.128	.	0.064	0.032	.	0.016
Indeterminate Sentence (Y/N)	0.393	.	0.182	0.080	.	0.037
Sentence Length (Months)	0.409	1.858	0.213	0.074	1.083	0.037
OGRS1 Score	0.105	0.941	0.054	0.011	0.992	0.030
OGRS2 Score	0.117	0.873	0.056	0.015	0.948	0.032
RoSH: High	0.312	.	0.115	0.015	.	0.005
RoSH: Low	-0.080	.	0.006	0.000	.	0.000
RoSH: Medium	-0.425	.	0.136	-0.029	.	0.009
RoSH: Very High	0.129	.	0.027	0.019	.	0.004
Complexity Score (0-5)	-0.045	1.261	0.050	0.007	1.221	0.040

Table J2

Balance Statistics Before and After Matching for IRiS Population

	Unmatched			Matched		
	Standardised Mean Difference	Variance Ratio	Kolmogorov-Smirnov Statistic	Standardised Mean Difference	Variance Ratio	Kolmogorov-Smirnov Statistic
Age (Years)	-0.127	1.103	0.138	0.029	1.153	0.053
Release Year	-0.309	1.050	0.136	-0.063	1.044	0.047
Ethnicity: Arab	-0.037	.	0.001	0.000	.	0.000
Ethnicity: Asian	-0.073	.	0.014	0.017	.	0.003
Ethnicity: Black	0.204	.	0.075	0.009	.	0.003
Ethnicity: Mixed	0.063	.	0.018	0.046	.	0.013
Ethnicity: Other	-0.052	.	0.003	0.000	.	0.000
Ethnicity: Refusal	0.184	.	0.036	-0.034	.	0.007
Ethnicity: White	-0.235	.	0.111	-0.028	.	0.013
Offence: Absconding/Bail Offences	-0.044	.	0.002	0.000	.	0.000
Offence: Burglary (Domestic)	-0.039	.	0.006	0.000	.	0.000
Offence: Burglary (Other)	0.049	.	0.006	0.000	.	0.000
Offence: Criminal Damage	-0.002	.	0.000	0.000	.	0.000
Offence: Drink Driving	0.088	.	0.010	0.000	.	0.000
Offence: Drug Import/Export/Production	0.065	.	0.008	0.000	.	0.000
Offence: Drug Possession/Supply	0.098	.	0.016	0.021	.	0.003
Offence: Fraud and Forgery	-0.044	.	0.002	0.000	.	0.000
Offence: Handling Stolen Goods	-0.023	.	0.001	0.000	.	0.000
Offence: Other Motoring	-0.007	.	0.001	0.000	.	0.000
Offence: Other Offence	-0.124	.	0.015	0.000	.	0.000
Offence: Public Order	-0.008	.	0.002	0.040	.	0.010
Offence: Robbery	-0.320	.	0.087	-0.049	.	0.013
Offence: Sexual (Against Child)	-0.254	.	0.050	-0.017	.	0.003
Offence: Sexual (Not Against Child)	-0.014	.	0.004	0.027	.	0.007
Offence: Soliciting or Prostitution	-0.017	.	0.000	0.000	.	0.000
Offence: Taking and Driving Away	0.075	.	0.009	0.000	.	0.000
Offence: Theft (Non-Motor)	-0.143	.	0.020	0.000	.	0.000
Offence: Theft from Vehicles	0.114	.	0.013	0.087	.	0.010
Offence: Violence	0.259	.	0.128	-0.027	.	0.013
Indeterminate Sentence (Y/N)	-0.456	.	0.089	0.034	.	0.007
Sentence Length (Months)	-0.421	0.358	0.093	0.042	1.196	0.057
OGRS1 Score	0.283	1.019	0.167	0.023	1.135	0.053
OGRS2 Score	0.273	0.939	0.161	0.003	1.149	0.047
RoSH: High	0.228	.	0.089	0.017	.	0.007
RoSH: Low	0.065	.	0.008	-0.029	.	0.003
RoSH: Medium	-0.346	.	0.118	-0.049	.	0.017
RoSH: Very High	0.109	.	0.021	0.068	.	0.013
Complexity Score (0-5)	0.070	1.038	0.042	0.062	1.061	0.030

Table J3

Balance Statistics Before and After Matching for Intervention Population

	Unmatched			Matched		
	Standardised Mean Difference	Variance Ratio	Kolmogorov-Smirnov Statistic	Standardised Mean Difference	Variance Ratio	Kolmogorov-Smirnov Statistic
Age (Years)	0.105	0.949	0.075	0.068	1.007	0.043
Release Year	-0.066	0.837	0.074	0.047	0.965	0.028
Ethnicity: Arab	-0.039	.	0.001	0.000	.	0.000
Ethnicity: Asian	-0.132	.	0.023	-0.032	.	0.006
Ethnicity: Black	-0.044	.	0.012	-0.081	.	0.021
Ethnicity: Mixed	-0.107	.	0.024	-0.043	.	0.010
Ethnicity: Other	0.050	.	0.004	0.055	.	0.005
Ethnicity: Refusal	0.033	.	0.003	0.041	.	0.003
Ethnicity: White	0.141	.	0.053	0.076	.	0.029
Offence: Absconding/Bail Offences	-0.046	.	0.002	0.000	.	0.000
Offence: Burglary (Domestic)	-0.114	.	0.015	-0.017	.	0.002
Offence: Burglary (Other)	0.013	.	0.001	0.006	.	0.001
Offence: Criminal Damage	0.049	.	0.011	0.053	.	0.012
Offence: Drink Driving	-0.020	.	0.001	0.000	.	0.000
Offence: Drug Import/Export/Production	-0.021	.	0.001	0.000	.	0.000
Offence: Drug Possession/Supply	0.022	.	0.003	0.000	.	0.000
Offence: Fraud and Forgery	-0.046	.	0.002	0.000	.	0.000
Offence: Handling Stolen Goods	-0.024	.	0.001	0.000	.	0.000
Offence: Other Motoring	-0.145	.	0.010	-0.008	.	0.001
Offence: Other Offence	-0.064	.	0.006	0.012	.	0.001
Offence: Public Order	-0.016	.	0.004	0.016	.	0.004
Offence: Robbery	0.002	.	0.001	0.006	.	0.002
Offence: Sexual (Against Child)	-0.160	.	0.036	-0.047	.	0.011
Offence: Sexual (Not Against Child)	-0.004	.	0.001	-0.057	.	0.015
Offence: Soliciting or Prostitution	-0.017	.	0.000	0.000	.	0.000
Offence: Taking and Driving Away	-0.005	.	0.000	0.000	.	0.000
Offence: Theft (Non-Motor)	-0.159	.	0.013	0.000	.	0.000
Offence: Theft from Vehicles	0.042	.	0.002	0.047	.	0.002
Offence: Violence	0.150	.	0.075	0.012	.	0.006
Indeterminate Sentence (Y/N)	0.309	.	0.136	0.020	.	0.009
Sentence Length (Months)	0.324	1.697	0.175	0.028	1.015	0.045
OGRS1 Score	0.136	0.956	0.068	0.060	1.060	0.048
OGRS2 Score	0.144	0.886	0.067	0.053	1.002	0.049
RoSH: High	0.297	.	0.110	0.006	.	0.002
RoSH: Low	-0.077	.	0.004	0.000	.	0.000
RoSH: Medium	-0.411	.	0.133	-0.024	.	0.008
RoSH: Very High	0.126	.	0.026	0.027	.	0.006
Complexity Score (0-5)	-0.028	1.223	0.041	-0.057	1.245	0.047

Appendix K

Remaining Kaplan-Meier Survival Plots

Figures K1 to K7 below are the remaining Kaplan-Meier survival plots for the IIRMS, IRiS and Intervention groups not already presented in Chapter 3.

Figure K1

Survival to Recall for IIRMS Group

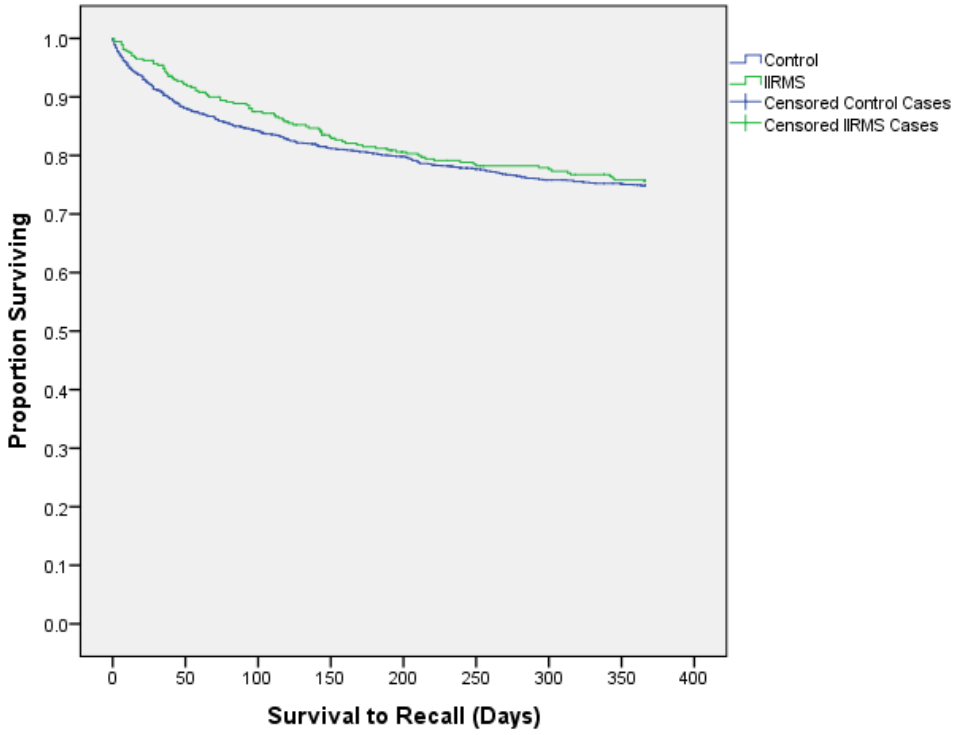


Figure K2

Survival to 100 Days for IIRMS Group

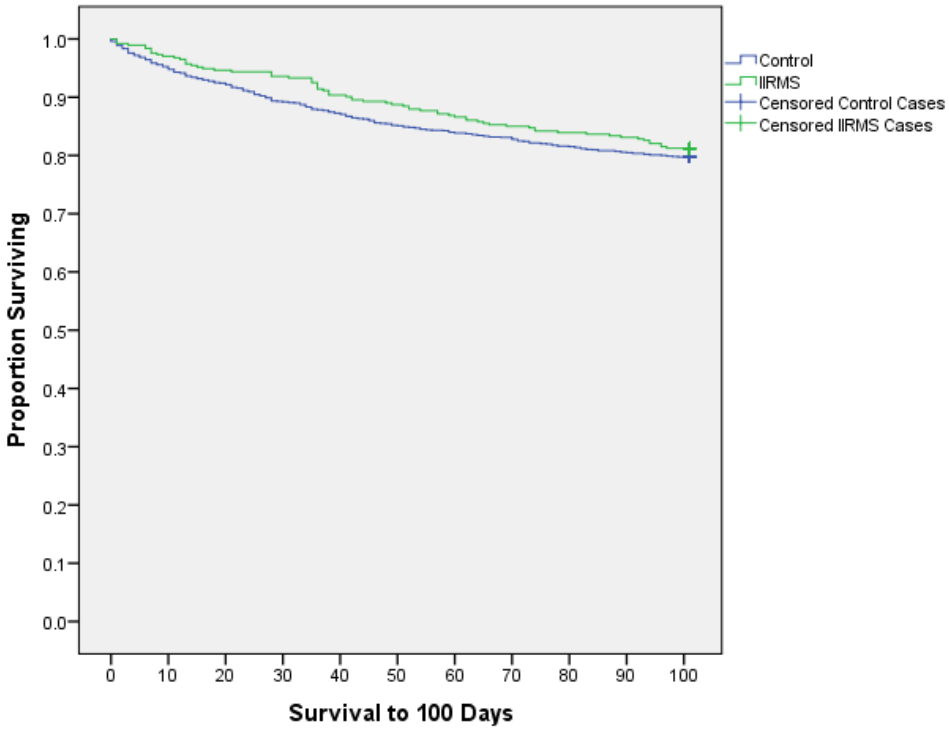


Figure K3

Survival to Recall for IRiS Group

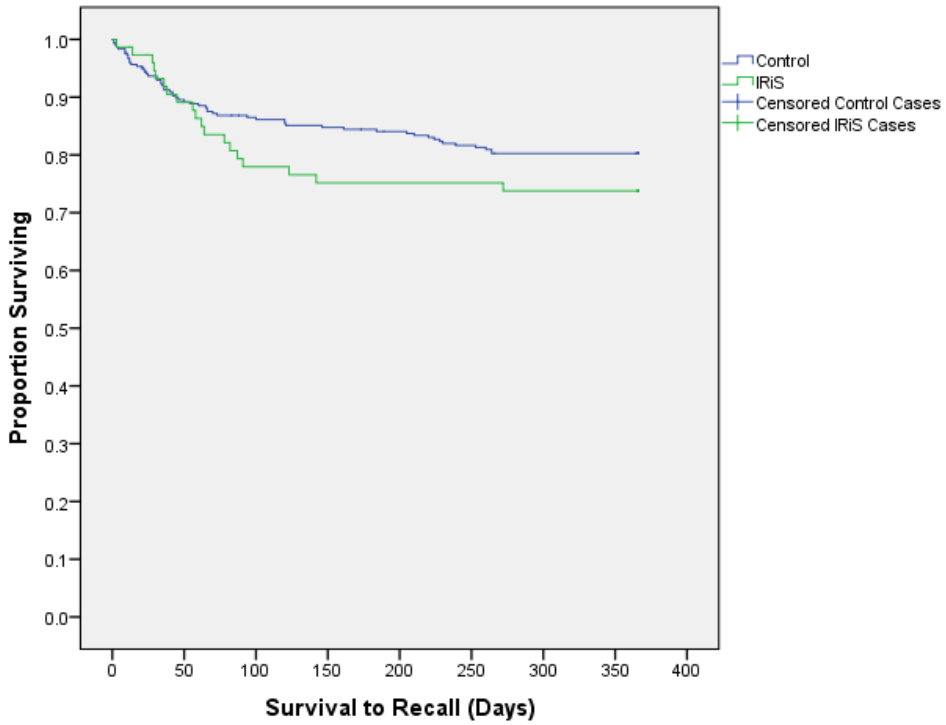


Figure K4

Survival to Reconviction for IRiS Group

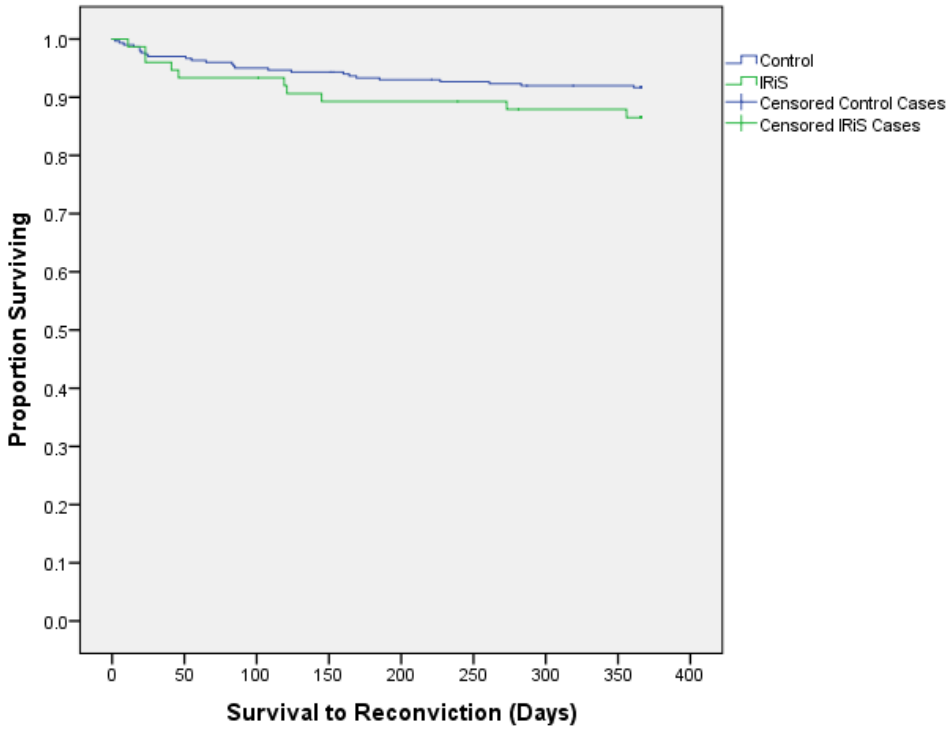


Figure K5

Survival to 100 Days for IRiS Group

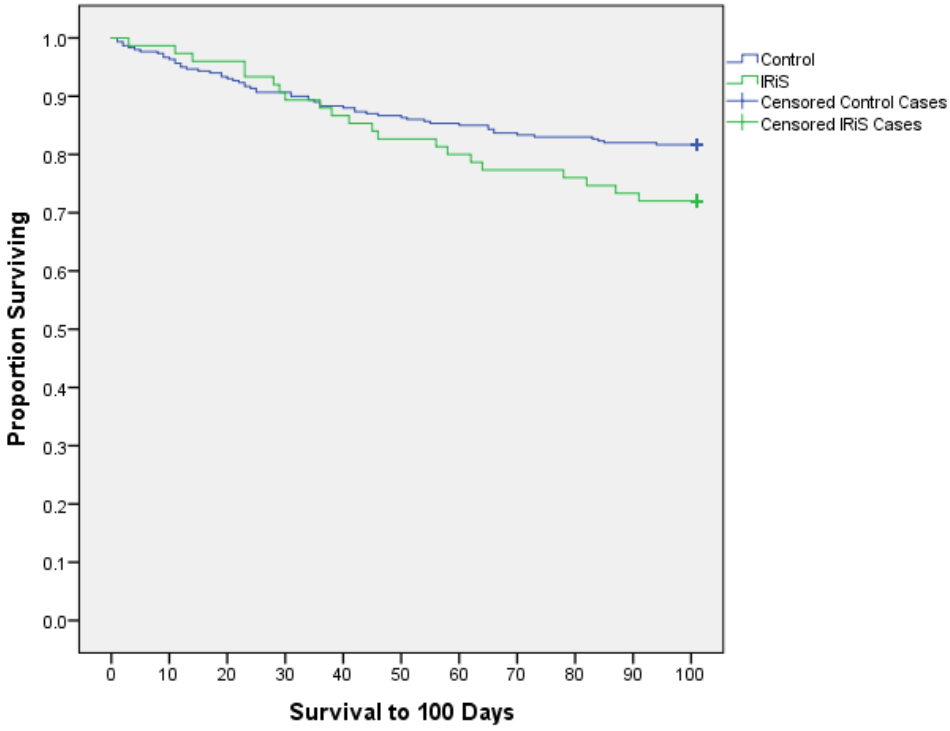


Figure K6

Survival to Recall for Intervention Group

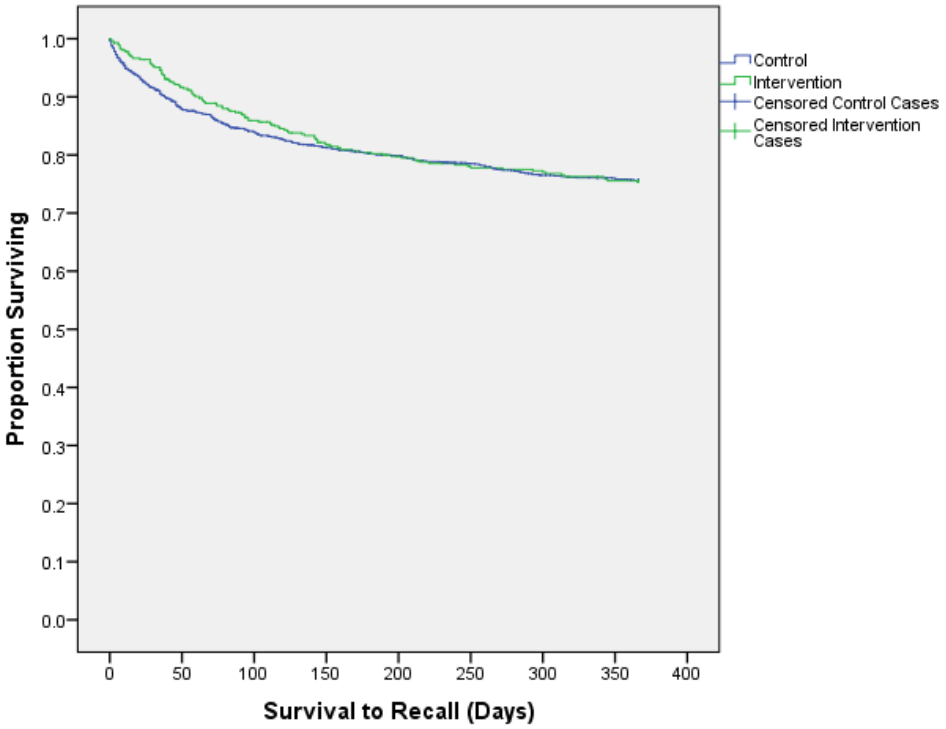
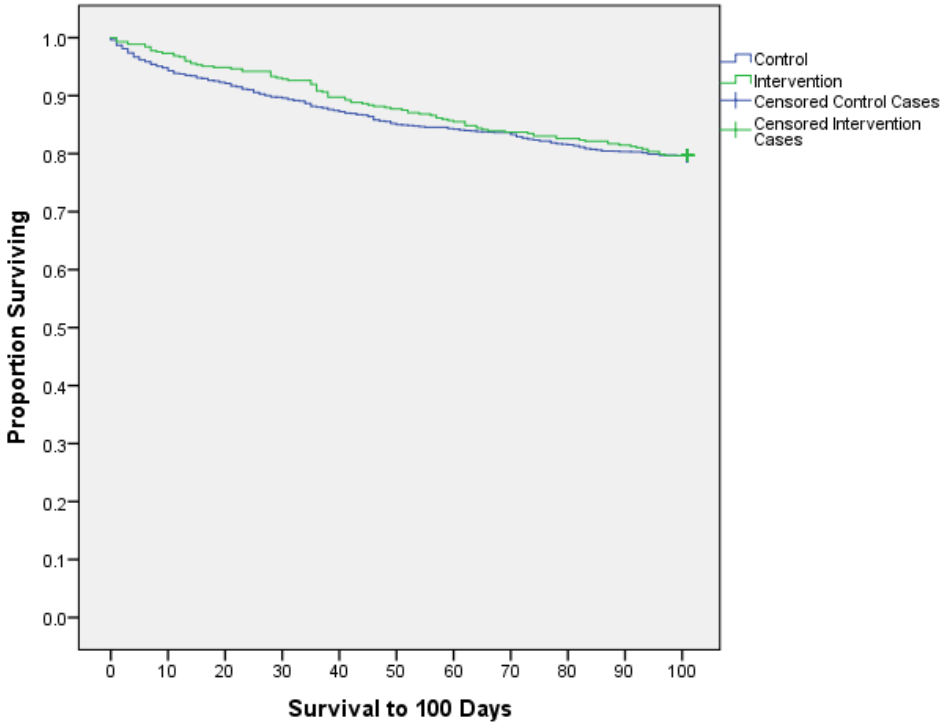


Figure K7

Survival to 100 Days for Intervention Group



Appendix L

Explanation of the Proportional Hazard Assumption Including Covariates Assessed for Time Dependence

Cox's Proportional Hazard Model requires the proportional hazard assumption to be met. This means that the hazard, or risk, of members of a population returning to custody needs to be constant across time. Analysis of Figures 5, 6 and 7 show that, in each case, the survival curves cross, indicating that the risk of returning to custody is not constant across time. As the proportional hazard assumption is not met, further exploration is required.

Table L1

Covariates Assessed for Time Dependence or Independence by Group

	IIRMS		IRiS		Intervention	
	Time Dependent	Independent of Time	Time Dependent	Independent of Time	Time Dependent	Independent of Time
Group (Control / Intervention)		X		X		X
Age		X		X		X
Ethnicity (White / Non-White)	X			X		X
Offence (Violent / Other)		X		X		X
Sentence Type (Determinate / Indeterminate)	X			X	X	
Sentence Length	X			X	X	
OGRS1 Score	X			X		X
OGRS2 Score	X			X		X
RoSH	X			X		X
OASys Items (Absent / Present)		X		X		X
Childhood Difficulties (Absent / Present)		X		X	X	
Mental Health (Absent / Present)		X		X		X
Self Harm / Suicide Attempts (Absent / Present)		X		X		X
Challenging Behaviour (Absent / Present)		X		X		X
Complexity Score (0-5)		X		X		X

Each covariate was examined separately across the IIRMS, IRiS and Intervention populations for their interaction with time. Table L1 indicates which covariates were time dependent within each group. The IIRMS population contained the most time dependent covariates, with six of the 15 covariates having an interaction with time. The IRiS population had no covariates with an interaction with time and the Intervention population contained three covariates with an

interaction with time. This indicates that the proportional hazard assumption is not met for some of the covariates in the IIRMS and Intervention populations, requiring the need to control for the effect of time on these covariates. This was done by including 'dummy' covariates accounting for the interaction with time for each of the time dependent covariates in the regression model.

Appendix M

CORE-OM Dimensional Breakdown

Table M1

Dimensional Breakdown of the CORE Outcome Measure (reproduced from Core System Handbook, 1998)

Dimension	Item	Severity / Intensity	Item No
Subjective Well Being	I have felt O.K. about myself. **Pos**	Lo	4
Subjective Well Being	I have felt like crying	Hi	14
Subjective Well Being	I have felt optimistic about my future **Pos**	Lo	17
Subjective Well Being	I have felt overwhelmed by my problems	Hi	31
Symptoms – anxiety	I have felt tense, anxious or nervous	Lo	2
Symptoms – anxiety	Tension and anxiety have prevented me doing important things	Hi	11
Symptoms – anxiety	I have felt panic or terror	Hi	15
Symptoms – anxiety	My problems have been impossible to put to one side	Lo	20
Symptoms – depression	I have felt totally lacking in energy and enthusiasm	Hi	5
Symptoms – depression	I have felt despairing or hopeless	Hi	23
Symptoms – depression	I have felt unhappy	Lo	27
Symptoms – depression	I have thought I am to blame for my problems and difficulties	Lo	30
Symptoms – physical	I have been troubled by aches, pains or other physical problems	Lo	8
Symptoms – physical	I have difficulty getting to sleep or staying asleep	Lo	18
Symptoms – trauma	I have been disturbed by unwanted thoughts and feelings	Hi	13
Symptoms – trauma	Unwanted images or memories have been distressing me	Hi	28
Functioning – general	I have felt able to cope when things go **Pos** Wrong	Hi	7
Functioning – general	I have been happy with the things I **Pos** have done	Lo	12
Functioning – general	I have been able to do most things I needed to **Pos**	Lo	21
Functioning – general	I have achieved the things I wanted to **Pos**	Hi	32

Functioning - close rel.	I have felt terribly alone and isolated	Hi	1
Functioning - close rel.	I have felt I have someone to turn to for support when needed **Pos**	Lo	3
Functioning - close rel.	I have felt warmth and affection for someone **Pos**	Lo	19
Functioning - close rel.	I have thought I have no friends	Hi	26
Functioning - social rel.	Talking to people has felt too much for me	Hi	10
Functioning - social rel.	I have felt criticised by other people	Lo	25
Functioning - social rel.	I have been irritable when with other people	Lo	29
Functioning - social rel.	I have felt humiliated or shamed by other people	Hi	33
Risk/Harm to self	I have thought of hurting myself	Lo	9
Risk/Harm to self	I have hurt myself physically or taken dangerous risks with my health	Hi	34
Risk/Harm to self	I made plans to end my life	Hi	16
Risk/Harm to self	I have thought it would be better if I were dead	Lo	24
Risk/Harm to others	I have been physically violent to others	Hi	6
Risk/Harm to others	I have threatened or intimidated another person	Hi	22

(Pos = Positively phrased item)