Violent reoffending in people released from prison: psychiatric epidemiology, risk assessment and psychological interventions

Gabrielle Beaudry

Wadham College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

Où est allé tout ce monde qui avait quelque chose à raconter?
On a mis quelqu'un au monde. On devrait peut-être l'écouter.

Un musicien parmi tant d'autres

Serge Fiori

Preface

The work described in this thesis was conducted under the supervision of Professor Seena Fazel and Dr Rongqin Yu between Michaelmas 2018 and Trinity 2022. It should be noted that all empirical chapters have been adapted from published material arising from research undertaken whilst reading for this degree. I confirm that this work is mine, and has been produced with no more input than would be considered appropriate or acceptable for a thesis. I have obtained written permission from my supervisors and other co-authors to include the material in my thesis. Detailed contribution statements are provided at the end of each relevant chapter. This thesis has not been submitted for any other degree at any other university.

This DPhil research was funded by the *Fonds de recherche du Québec—Santé* (FRQS; grants number 255906 and 282526). I also received minor grants from Green Templeton College, Wadham College, and the Violence, Abuse and Mental Health Network (VAMHN). The views expressed in this thesis are mine and not necessarily those of these organisations.

This thesis was written using the Oxford thesis LaTeX template first developed by Keith A. Gillow, and subsequently modified by Sam Evans and John McManigle.

References and appendices for all chapters are presented at the end of the thesis.

Word count (exclusive of bibliography, appendices, figures and tables): 36,646.

Publications arising from this thesis

Published journal articles

- **Beaudry, G.**, Yu, R., Alaei, A., Alaei, K., & Fazel, S. (2022). Predicting violent reoffending in individuals released from prison in a lower-middle-income country: a validation of OxRec in Tajikistan. *Frontiers in Psychiatry*, *13*(85141), 1–9.

- **Beaudry, G.**, Yu, R., Perry, A., & Fazel, S. (2021). Effectiveness of psychological interventions in prison to reduce recidivism: a systematic review and meta-analysis of randomised controlled trials. *Lancet Psychiatry*, *8*(9), 759–773.

- **Beaudry, G.**, Yu, R., Långström, N., & Fazel, S. (2020). An updated systematic review and metaregression analysis: mental disorders among adolescents in juvenile detention and correctional facilities. *Journal of the American Academy of Child and Adolescent Psychiatry*, *60*(1), 46–60.

Journal articles in preparation

- **Beaudry, G.**, Yu, R., Prescott-Mayling, L., Miller, O., Fanshawe, T.R., & Fazel, S. (in preparation). Prediction of violent reoffending in people released from prison in England: external validation study of a risk assessment tool (OxRec).

Acknowledgements

I would like to express my deepest gratitude to those who have made this work possible. This thesis is as much mine as it is theirs.

To my supervisors, Professor Seena Fazel and Dr Rongqin Yu. Seena, for challenging me to think critically as an independent researcher, and embodying what it means to conduct research that is timely, impactful and clinically relevant. Rongqin, for always making time, and for showing me that intellectual rigour and empathy can co-exist in academia. Thank you to Mrs Diana Garrod for her invaluable administrative support.

To my colleagues, Danny, Denis and Howard for their advice along the way. Shaoling and Verena, from whom I have learned so much. Julia and Shannon, for sharing the ups and downs of this research journey with me. It has been a privilege to work with and alongside you, and I look forward to many more collaborations in the future.

To my friends, on both sides of the pond (Montréal and Oxford), for supporting me near and far. Especially Anne, without whom this endeavour would not have made sense at times. My friends from GTC—Alice, Victoria, Trishna, Minh, Andreas, Tom, and all the others who have made my time there so enjoyable. Claire, qui comprend si bien ce qu'est d'être une franco, un peu trop anglo. Antje and Jon, for sharing my love of great food, coffee and long runs. Lea, da Leicester a Sirmione. Phil, the best lab partner turned friend one could ask for. Ericka, qui trouve toujours les bons mots. Julia, for simply being the best teammate on and off the court. Julien, for Wadham—and all the rest. Catherine, encore et toujours. Mariane, Nicolas, Marion et les Juliens, for the past decade and those yet to come. Marie-Joëlle et Mathias, pour cette amitié qui transcende les frontières. Merci de m'inspirer à rêver plus grand et plus loin.

To my partner, Fabian, for making everywhere we go feel like home.

À ma famille, pour tout. Ma soeur, Emma, fidèle complice qui rend ma vie plus lumineuse. Mes parents, Nathalie et René, pour leur amour et leur soutien inébranlable. Merci de m'avoir appris, sans un mot, à ne jamais abandonner. Mes grands-parents, Liliane, Marthe, Roch et Jacques, qui ont semé les graines de cette aventure il y a de cela plusieurs décennies.

Abstract

Violence was identified as a global public health concern by the World Health Assembly nearly three decades ago. Despite reported decreases in violent crime in many countries, reoffending rates worldwide remain high. Amongst people released from prison, there are some at high-risk of perpetrating interpersonal violence. Identifying these key individuals, who are most in need of effective interventions to prevent future criminality, is crucial to reducing societal violence, as their contribution to this major problem is considerable. In this thesis, I focus on violence risk assessment and prevention of future violence in people released from prison by employing methods from psychiatric epidemiology, public mental health and prediction modelling. I start by estimating the prevalence of a modifiable risk factor for violence (i.e. treatable mental disorders) amongst adolescents in juvenile detention and correctional facilities. I select this subgroup of the global prison population as most severe mental disorders emerge in late adolescence, and thus this period provides a critical window to improve prognosis and intervention. My second and third studies externally validate a novel, scalable and transparent violence prediction model—the Oxford Risk of Recidivism (OxRec) tool—in two new countries. I investigate the predictive ability of OxRec in both lower middle-income and high-income settings using data from Tajikistan and England to identify individuals who could be targeted for empirically supported interventions in prison and on release. Lastly, I evaluate the effectiveness of widely implemented psychological interventions for people in prison to reduce offending after release. I synthesise the evidence by solely including randomised controlled trials to identify the current most effective treatments, and inform future evidence-based research and policy in this area.

Contents

List of Figures

List of Tables

List of Abbreviations

ADHD . . . . Attention-Deficit Hyperactivity Disorder.

AUC . . . . . . Area Under the Curve.

CDC . . . . . . Centers for Disease Control and Prevention.

COMPAS . . . Correctional Offender Management Profiling for Alternative Sanctions.

DALY . . . . . Disability-Adjusted Life Year.

GHQ . . . . . . General Health Questionnaire.

HIC . . . . . . High-Income Country.

HIV . . . . . . Human Immunodeficiency Virus.

LMIC . . . . . Low- and Middle-Income Country.

MOOSE . . . . Meta-analysis Of Observational Studies in Epidemiology.

NOMS . . . . National Offender Management Service.

NPV . . . . . . Negative Predictive Value.

OASys . . . . . Offender Assessment System.

OxRec . . . . . Oxford Risk of Recidivism.

PAF . . . . . . Population Attributable Risk Fractions.

PPV . . . . . . Positive Predictive Value.

PTSD . . . . . Posttraumatic Stress Disorder.

PROBAST . . Prediction model Risk Of Bias ASsessment Tool.

PRISMA . . . Preferred Reporting Items for Systematic Review and Meta-Analyses.

RCT . . . . . . Randomised Controlled Trial.

ROC . . . . . . Receiver Operating Characteristic Curve.

RNR . . . . . . Risk, Need and Responsivity.

SRQ . . . . . . Self-Reported Questionnaire.

SUD . . . . . . Substance Use Disorder.

TC . . . . Therapeutic Community.

TRIPOD . . . Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis.

UI . . . . Uncertainty Intervals.

UK . . . . United Kingdom (British).

UNICEF . . . United Nations Children's Emergency Fund.

US . . . . United States (American).

WHO . . . . . World Health Organisation.

1
Introduction

Contents

Violence is a major and growing public health problem which claims the lives of more than a million every year (Butchart et al., 2014; World Health Assembly, 1996; World Health Organization, 2014a). According to the most recent estimates from the World Health Organization (WHO), approximately 1.3 million individuals die annually as a result of violence, accounting for 2.5% of all deaths worldwide (World Health Organization, 2014a). The impact of violence extends well beyond these deaths when considering the additional non-fatal consequences of being the victim or witness to violent acts (Krug, Mercy, et al., 2002). Those not only include physical injuries, but also a myriad of adverse behavioural, cognitive, mental, sexual and reproductive health outcomes, as well as chronic diseases (World Health Organization, 2014a). This largely predictable and preventable public health issue also imposes significant social and economic costs on governments, which translate in large expenditures of public funds for legal, social and health care services (Decker et al., 2018). The financial burden is disproportionately large on low- and middle-income countries (LMICs), which account for approximately 90% of the global violence-related deaths (Krug, Dahlberg, et al., 2002; Waters et al., 2005).

Over the past 20 years, the world prison population has risen by 24%—with nearly 11 million people currently living in incarceration (Fair & Walmsley, 2021), and 30 million being released from prison in any given year (Schmitt & Warner, 2010). Amongst people released from prison, there is a group at high risk of perpetrating interpersonal violence (Fazel, Chang, et al., 2016). Yet, current approaches to violence risk assessment and management in this key population are partly limited. This is reflected by high recidivism rates in most countries, which have not decreased significantly in recent decades (Fazel & Wolf, 2015; Yukhnenko et al., 2020). This thesis will examine violent reoffending in people released from prison to address important gaps in the epidemiological and prediction modelling literature.

This chapter summarises the background literature to identify these knowledge gaps. It begins with a brief historical context of violence as a public health concern, defines violence with a particular focus on violent reoffending, and then reviews the relevant research on the prediction of recidivism. I highlight three specific areas for conducting new research: (1) mental health problems in one high risk group in prison, adolescents; (2) testing the validity of a novel risk assessment tool for violent reoffending in people in prison; and (3) the trial evidence for psychological treatments aimed at reducing recidivism. I conclude the chapter with a structural overview of the thesis and highlight implications of this work for methods and practice, alongside new research questions.

Person-centered language is used throughout this thesis to describe individuals who experience the criminal justice system, their characteristics, and experiences. The use of respectful and appropriate language when researching this key population (e.g. 'people in prison' in place of 'prisoners') can contribute to improving access to health services, minimise discrimination, and reduce barriers to successful community reintegration (Tran et al., 2018).

1.1 Brief historical context

What constitutes violence, and violence-prevention programs and policies have significantly changed over time, reflecting wider shifts in social and political movements (Pache, 2020). The notion that violence is a public health issue that can be effectively prevented is a relatively recent development, dating only from the past four decades (Krug, Mercy, et al., 2002; Niolon et al., 2020). Several important trends and historical developments contributed to violence being recognised and acknowledged as a matter of public health, first at a national-level and subsequently on a global scale.

In 1979, a report from the Surgeon General of the United States introduced violence as one of the 15 health priority areas for the country (US Department of Health & Welfare, 1979). Key factors such as (1) therapeutic advances in infectious diseases; (2) increased rates of intentional injuries amongst specific priority groups (e.g. young African Americans, particularly male individuals) in the 1980s; (3) growing awareness of the role of behavioural factors in disease prevention; and, (4) emergence of child maltreatment and intimate partner violence as social and medical problems (beyond the criminal justice system) in the 1960–1970s, sparked state and federal interest in violence reduction (Dahlberg & Mercy, 2009; Niolon et al., 2020). This culminated in the establishment of the Violence Epidemiology Branch within the Centers for Disease Control and Prevention (CDC) in 1983, and ultimately in 1992, the creation of an independent governmental agency to centralise research and policy efforts in violence prevention (i.e. the National Center for Injury Prevention and Control) (Rosenberg, 1985; Sleet et al., 2012).

On the international front, violence was legitimised as a global public health problem in 1996 when the Forty-Ninth World Health Assembly adopted Resolution WHA49.25. In this resolution, the Assembly called attention to the grave repercussions of violence at local, national, and international levels, and emphasised its detrimental effects on health care systems across the world. The Assembly also urged for a concerted action from member states and the Director-General of the WHO to develop evidence-based solutions for violence (World Health Assembly, 1996). This call for action resulted in the release of the *World report on violence and health* by the WHO in 2002—the first comprehensive review of the magnitude and impact of specific types of violence, and approaches to violence prevention and management (Krug, Mercy, et al., 2002). Altogether, these events have driven a step-change in violence prevention, whereby violence is now understood as part of the global public health agenda (Pache, 2020).

1.2 Defining violence

Violence can be defined in a variety of ways, and an international standard is still lacking. However, the concept of intentionality remains central to most definitions. In this thesis, I will adopt the WHO's definition of violence (from the *World report on violence and health*), as it is sufficiently general to encompass a broad array of differences between countries, cultures, and beliefs systems, and thus provides an international perspective on this complex phenomenon. Henceforth, violence is defined as "*the intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community, that either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment or deprivation*" (Krug, Dahlberg, et al., 2002, p.5).

1.2.1 Types of violence

According to this definition, violence can be divided in three broad categories based on the identity of the perpetrator and that of the victim; self-directed, interpersonal or collective violence (see Table 1.1 for specific definitions). Research has shown

Table 1.1: Definitions of the types of violence

	Definition
Self-directed	Any intentional act that may cause injury to self, including death.
Interpersonal	Intentional use of physical force or authority against another person, which can take the form of physical, sexual, or psychological abuse, as well as deprivation and neglect.
Collective	Instrumental use of violence by persons who identify as members of a group—whether temporary or permanent—against another group or collection of individuals in order to fulfill political, economic, or social ambitions.

Note. Definitions adapted from Krug, Dahlberg, et al., 2002; Mercy et al., 2017.

that particular types of violence also have unique aetiologies, and distinct physical and psychological sequelae, despite their many interconnections (Krug, Mercy, et al., 2002). For this reason, I will solely focus on interpersonal violence in this thesis.

The social-ecological model (Bronfenbrenner, 1979; Garbarino, 1985; Tolan & Guerra, 1994)—whereby violence is conceptualised by the dynamic interplay between various levels of influence on behaviour (i.e. individual, relationship, community and societal factors)—is the theoretical framework used by the WHO for understanding violence (Krug, Dahlberg, et al., 2002). This approach draws on evidence from several fields of research to advance the knowledge of the context, causes and consequences of violence in the lives of individuals, as well as the communities, and the broader socio-cultural and political situations in which they exist (Kelly, 2011). Understanding the links between specific types of violence through the lens of the ecological model, suggest that common risk factors may serve as targets to prevent simultaneously more than one type of violence (Krug, Dahlberg, et al., 2002; Krug, Mercy, et al., 2002).

1.2.2 Interpersonal violence

Interpersonal violence refers to violence that occurs between individuals, whether they be intimate partners, family members, friends, acquaintances or even strangers. It encompasses multiple forms of violence, including child maltreatment, youth violence, intimate partner violence, sexual violence, abuse of the elderly, random acts of violence perpetrated by strangers, and those committed in institutional settings (e.g. workplaces, health care facilities and prisons) (World Health Organization, 2014a). This type of violence significantly impacts both individuals (namely victims and perpetrators) and communities, with a wide range of adverse health effects that can persist over a lifetime and that often transcend generations (Widom & Wilson, 2015; World Health Organization, 2014a).

Measuring interpersonal violence

There are many ways to measure the health burden of interpersonal violence. Summary measures of premature mortality and morbidity (i.e. frequencies, rates and years of life lost [YLL]) are amongst the most common, but disability-adjusted life years (DALYs)—a measure of overall disease burden, expressed by the number of years impaired by disabilities or lost due to premature death—are also increasingly employed (Anand & Hanson, 1997; Macdonald, 2002). As per the *Global Burden of Disease* study published in *The Lancet*, an estimated 415,000 people worldwide died from interpersonal violence in 2019, making it the fifth leading cause of DALYs amongst adolescents and young adults (10–24 years) (Vos et al., 2020). However, mortality estimates should be interpreted with caution, as cross-national homicide research is limited by lacking or incomplete data from less affluent nations (Kanis et al., 2017). Non-fatal consequences of violence are more difficult to measure than fatalities, as many violence-related injuries are treated outside the health care system, whilst others remain untreated, or are simply never reported (World Health Organization, 2014b). In the United States (US), for example, hospital emergency departments alone recorded more than 1.5 million assault-related injuries in 2019, representing a ratio of 79:1 to homicides that year (Centers for Disease Control and Prevention, 2022).

Aside from the immediate physical injuries, interpersonal violence has also been linked to a variety of long-term health consequences, such as depression, substance misuse, and disabilities (World Health Organization, 2014b). Although comprehensive statistics for these other health repercussions are yet to be compiled, evidence on the economic impact of societal violence suggest that they constitute the largest share of the social and health burden arising from this serious issue (Waters et al., 2005; Waters et al., 2004).

Risk factors for interpersonal violence

Over the years, a substantial body of research has examined individual-level risk factors for interpersonal violence in longitudinal cohorts and population-based studies (Wolf et al., 2014). A recent umbrella review of 22 meta-analyses (Fazel et al., 2018), which included more than 120,000 participants from 1,139 individual studies across 14 different countries, found that neuropsychiatric disorders were amongst the strongest modifiable risk factors, in both relative and absolute terms. More specifically, substance use disorders (SUDs) presented the largest effect at the population level, as measured by population attributable risk fractions (PAF = 14.8%, 95% CI 9.0–21.6%). Personality disorders, in particular antisocial personality disorder, and schizophrenia-spectrum disorders were also related to violence. As for historical predictors, violence exposure in childhood was identified as the most important one (PAF = 12.2%, 95% CI 6.5–17.4%). Other established risk factors include various childhood (Derzon, 2001; H. W. Wilson et al., 2009), adolescence (Moffitt, 1993) and parental factors (e.g. youth antisocial behaviour and parental incarceration) (Murray et al., 2012), and sociodemographic characteristics such as younger age and male sex (Moffitt et al., 2001).

Far less attention has been paid to the social and economic determinants of interpersonal violence. Nonetheless, existing meta-analytic research has highlighted the importance of some neighbourhood and country-level factors, such as alcohol consumption, income inequality and urbanicity (Wolf et al., 2014). These will not be further discussed here, as they are not directly relevant to the overarching theme of this thesis—key challenges for people in and released from prison at the intersection of health and justice.

1.3 General and violent recidivism

Findings from governmental reports suggest that interpersonal violence has been on the decline in recent decades (Federal Bureau of Investigation (US), 2020; Office for National Statistics (UK), 2021), but national reoffending rates have not mirrored this downward trend, particularly in high-income countries (Fazel & Wolf, 2015; Yukhnenko et al., 2020). Reducing reoffending is a pressing challenge for societies, and wider social and economic costs of reoffending in the UK have been estimated at more than £18 billion per year (Newton et al., 2019). Moreover, one single-state study from the US found that each recidivism event engendered financial costs of approximately $150,000 (Steinfeld et al., 2018).

Reoffending, otherwise known as recidivism or repeat offending, refers to the recurrence of criminal activity following a period of incarceration or community-based sentence for a prior crime. Conversely, violent reoffending (or violent recidivism) is denoted by a return to violent crime. Whilst these definitions are broadly agreed upon, the measurement and operationalisation of reoffending outcomes vary considerably between and even within jurisdictions. International comparisons are complicated by countries reporting different recidivism statistics (e.g. rearrest, reconviction or reimprisonment for a new sentence or the breach of parole conditions), and varying follow-up periods that typically range from one to five years (Yukhnenko et al., 2020). Recidivism is often measured by rearrest, although arrest does not imply conviction, nor guilt (W. Sawyer & Wagner, 2022). Arrest is also an imprecise proxy for crime being as interaction with law enforcement may vary by neighbourhood, and therefore it likely contains systematic biases (Corbett-Davies et al., 2017). Rather than rearrest, other informative metrics include “conviction for a new crime, reincarceration, or a new sentence of imprisonment,” and these provide more accurate information on recidivism (W. Sawyer & Wagner, 2022). Further, the laws of a particular jurisdiction determine what is to be considered a violent crime, and thus some specific offences do not translate as being ‘violent’ from one province, state or country to another.

For the purpose of this thesis, violent reoffending is defined according to the Swedish criminal code, as understood by a new conviction for any violent crime. In Sweden, violent crime entails both the use (i.e. homicide, assault, robbery, arson, any sexual offence; rape, sexual coercion, child molestation, indecent exposure, or sexual harassment) or threatened use of violence against another person (i.e. illegal threats or intimidation). The rationale for this outcome definition is that it was previously used to develop OxRec (the Oxford Risk of Recidivism tool) with national population-based registers from Sweden (Fazel, Chang, et al., 2016), and this is the model that is being tested in this thesis amongst people released from prison in Tajikistan (Chapter 3) and England (Chapter 4). Some adaptations were made to the violent reoffending outcome and its operationalisation to enable comparison between countries, and those will be detailed in subsequent empirical chapters.

1.3.1 Recidivism rates

The rate at which people with convictions relapse into criminal behaviour is one of many important indicators to evaluate and compare the effectiveness of criminal justice systems worldwide. Recidivism rates vary around the globe, and estimates are based on different data sources, outcome definitions and time periods. The highest levels are observed in HICs, both in high-incarceration nations (e.g. the US) and in low-incarceration nations (e.g. Norway), where 50% to 70% of formerly incarcerated people are rearrested during the first two or three years following release (Durose et al., 2014; Nygaard Andersen & Skardhamar, 2017; Nygaard Andersen & Telle, 2022; Yukhnenko et al., 2019). There is a paucity of recidivism rate estimates from LMICs, but the actual number is likely to be much higher than in HICs due to high rates of serious crime and incarceration (Fair & Walmsley, 2021; World Health Organization, 2014a). More than 70% of the global population of imprisoned persons are based in LMICs (Fair & Walmsley, 2021), yet research in this key population is scarce (Ako et al., 2020; Gureje & Abdulmalik, 2019), and policymakers are often reluctant to invest in recidivism prevention and social reintegration programmes (Chin & Dandurand, 2018).

Global estimates of the rate of violent repeat offending are lacking, with few countries reporting this information in a comprehensive manner. However, available data from the US indicate that more than 20% of people released from prison are arrested for a violent crime during the first three years of the follow-up period (Alper et al., 2018). People released from prison account for nearly one-fifth of all crimes committed in any given year in the US, although most of these reoffending events are non-violent (Petersilia, 2011). Individuals serving prison sentences typically have higher recidivism rates than those receiving community sentences, despite the fact that the two populations have different baseline characteristics. A recent meta-analysis of 28 cohort studies of individuals receiving community or suspended sentences in 19 countries found that one-year reoffending rates varied between 5% and 33%, and that two-year rates ranged from 16% to 41% (Yukhnenko et al., 2020). By contrast, another meta-analysis of 28 cohort studies estimated an average global incidence between 26% and 60% (for rearrest), 20% to 63% (for reconviction), and 14% to 45% (for reimprisonment) within two years of release from prison (Fazel & Wolf, 2015; Yukhnenko et al., 2019).

1.3.2 People who experience incarceration

More than 11.5 million individuals are currently incarcerated around the world (Fair & Walmsley, 2021), and another 30 million circulate through prisons per year (Schmitt & Warner, 2010). Prison population rates vary greatly between and within continents, but developing nations account for almost two-thirds of the total prison population. The largest number of people in prison is found in the US, with more than two million individuals and the highest prison population rate at 629 per 100,000 in contrast with a median of 140 per 100,000 worldwide. Several other countries also have a total number which exceeds the quarter of a million, being China (1.69 million), Brazil (811,000), India (478,000), Russia (471,000), Thailand (309,000), Turkey (291,000) and Indonesia (266,000) (Fair & Walmsley, 2021). Many children and adolescents experience incarceration, although similar comprehensive figures have not been released due to a lack of reliable data from most countries and the large variety of facilities (e.g. youth-specific institutions, or adult prisons). However, the United Nations Children's Fund (UNICEF) has estimated that over one million children are held in prisons and other closed facilities in response to crime, but primarily for acts that are not inherently criminal, such as immigration offences (Bochenek, 2017).

Complex health and social care needs

Adolescence is a critical phase for several health and developmental conditions that are important antecedents of criminal justice involvement beyond youth crime (Hughes et al., 2020; Murray & Atilola, 2020). Adults and young people who experience incarceration are affected by a wide array of complex care needs, including mental health and substance use issues, multimorbidity, poverty and social marginalisation. Rates of physical and psychiatric morbidity are higher than the general population, as are suicide and all-cause mortality rates, both within prisons and on release (Fazel & Baillargeon, 2011; Fazel, Hayes, et al., 2016; Fazel, Ramesh, et al., 2017; Kinner et al., 2013; Kinner & Young, 2018; Willoughby et al., 2021; Zlodre & Fazel, 2012).

Premature mortality

A considerable proportion of premature deaths in prison populations are caused by suicide and violence, and could therefore be avoided. Detailed examination of the prevalence of suicide in prisons by Fazel, Ramesh, et al. (2017) showed that prison suicide rates in HICs range from 23 to 100 suicides per 100,000 people in prison, with the lowest rates found in Australasian and North American nations, and the highest rates recorded in Nordic countries. For comparison, the WHO estimates that the global annual suicide rate is 10.7 per 100,000 individuals, with non-negligible differences between age groups and nations (Bachmann, 2018; WHO, 2022). Deaths by suicide in prison are the result of a complex interplay of contributing factors, the most important of which are individual-level factors (Fazel, Ramesh, et al., 2017). Targets for preventive efforts have been identified by means of research on potentially modifiable risk factors including clinical correlates (e.g. suicidal ideation, self-harm history, and more generally unmet mental health needs) and institutional ones (i.e. single-cell occupancy and lack of visits) (Zhong, Senior, et al., 2021).

The prevalence of self-harm is also higher amongst people who experience incarceration compared with their community-residing counterparts (5–6% and 20–24% in male and female individuals, respectively vs. <1%) (Borges et al., 2010; Hawton et al., 2014; Jenkins et al., 2005; Klonsky, 2011), and those who self-harm have a six to eightfold greater risk of suicide whilst being incarcerated and following release (Fazel, Cartwright, et al., 2008; Haglund et al., 2014; Humber et al., 2013; Pratt et al., 2010). Some underlying factors of this leading cause of morbidity overlap with those from suicide (i.e. current suicidal ideation and solitary confinement), but others appear to be specific to self-harm (e.g. homelessness, being sentenced for five years or more and childhood sexual abuse) (Favril et al., 2020).

Individuals who experience incarceration have a greater risk of violence victimisation before, during, and after serving their prison sentence, compared to those who have never been incarcerated (Jennings et al., 2012; Pérez et al., 2010; N. Wolff et al., 2007; N. Wolff et al., 2009a, 2009b). A group of Australian researchers has suggested that this may be partly due to violence victimisation and incarceration sharing several underlying factors, such as psychiatric disorders, substance misuse, homelessness, and financial precariousness. Importantly, this risk also translates to that of violence-related mortality after release from incarceration. As such, people released from prison are nearly eight times more likely to die from violence than their general population counterparts (Willoughby et al., 2021).

Psychiatric morbidity

There is extensive evidence of the high prevalence of mental illnesses in this specific population, yet access to adequate psychiatric care in correctional facilities remains problematic (Fovet et al., 2022). However, Fazel, Hayes, et al. (2016) recommend caution in the interpretation of prevalence findings because of several caveats to primary studies and meta-analytic methodology. For instance, surveys of specific disorders (i.e. attention-deficit hyperactivity disorder [ADHD]) (Gaïffas et al., 2014; Usher et al., 2013; Young et al., 2015), and those relying on approaches to clinical diagnosis other than professional judgement (i.e. self-report or laypersons) are particularly prone to prevalence overestimation (Fazel & Danesh, 2002). The quality of individual studies is often highly variable, and thus some meta-analyses combine small and selected samples that likely contain extreme values. The choice of statistical model when pooling prevalence estimates also plays a role; in general, random-effects yield more conservative summaries and generate wider confidence intervals than fixed-effects (Poole & Greenland, 1999). Despite these limitations, high-quality systematic reviews and meta-analyses were conducted in recent years to synthesise the rapidly growing medical literature, and provide reliable global and regional estimates about treatable mental disorders in people who experience incarceration (Baranyi et al., 2018; Baranyi et al., 2019; Facer-Irwin et al., 2019; Fazel & Baillargeon, 2011; Fazel, Hayes, et al., 2016; Fazel & Seewald, 2012; Fazel et al., 2018; Fazel, Yoon, et al., 2017).

Fazel and Seewald (2012) reviewed 109 prevalence studies for severe mental illness (schizophrenia-spectrum disorders and major depression) including 33,588 people in prison from 24 countries. The authors found no significant sex-related differences, as pooled prevalence estimates were around 3.6% (95% CI 3.1–4.2) in male individuals and 3.9% (95% CI 2.7–5.0) in female individuals for psychosis, and 10.2% (95% CI 8.8–11.7) in male individuals and 14.1% (95% CI 10.2–18.1) in female individuals for depression. However, the prevalence rates of psychosis were significantly higher in LMICs than in HICs (5.5% vs. 3.5%). Moreover, there was no evidence of increasing trends over time, with the exception of those observed in prevalence studies of depression from the US (from 1974–2005) (Fazel & Seewald, 2012).

Another review also focused on severe mental illness (Baranyi et al., 2019), but searched specifically for prevalence studies conducted in LMICs, as few studies from less-affluent nations were identified in the previous one, and epidemiological evidence on mental illness in prisons from LMICs is generally lacking (Jack et al., 2018). Consistent with findings from HICs, there was evidence of high prevalence of psychotic disorders (6.2%, 95% CI 4.0–8.6) and major depression (16.0%, 95%

CI 11.7–20.8), with elevated rates at arrival to prison. On the one hand, general population reviews suggest that the prevalence of schizophrenia-spectrum disorders and major depression is lower in LMICs than in HICs. On the other hand, Baranyi et al. (2019) found no evidence of such difference amongst people in prison, which may be explained by barriers to accessing mental health care for incarcerated populations in resource-poor settings (Jack et al., 2018).

Posttraumatic stress disorder (PTSD) is also widespread in incarcerated populations due to disproportionate levels of lifetime exposure to various forms of violence (emotional, sexual and physical) and other traumatic events. People who experience incarceration have extensive histories of trauma (Ardino, 2011, 2012; Garbarino, 1995), and traumatic exposure (e.g. child maltreatment, sexual abuse and witnessing violence) has previously been linked to antisocial behaviour (Ardino et al., 2013; Dziuba-Leatherman & Finkelhor, 1994). Approximately 6.2% (95% CI 3.9–9.0) of male and 21.1% (95% CI 16.9–25.6) of female individuals were reported to have PTSD in a recent review, based on point prevalence estimates (Baranyi et al., 2018). These findings underscore significant sex differences in prevalence rates, and suggest that the point prevalence of PTSD is nearly five times greater in male individuals and eight times higher in female individuals in prison compared with the general population (Kessler et al., 1995; Stein et al., 1997).

Alcohol and drug use disorders are often associated with the aforementioned psychiatric disorders, and some research suggests that substance misuse accounts for a large proportion of mortality following release from prison (Binswanger et al., 2013; Chang et al., 2015). People who inject drugs also have an increased risk of being infected with blood-borne diseases and spreading them to others, yet access to pharmacotherapies in prison (e.g. methadone maintenance treatment) is extremely limited (Chandler et al., 2009). Fazel, Yoon, et al. (2017) updated the evidence on prevalence studies of SUDs amongst people in prison by building on previous work from Fazel et al. (2006). The authors found that SUDs were highly prevalent in this population, and identified substantial heterogeneity arising from clinical and methodological differences between individual studies. About one-fourth of all newly incarcerated individuals had an alcohol use disorder (24.0%, 95% CI 21.0–27.0). Prevalence of drug use disorder differed significantly by sex, with pooled estimates of 51.0% (95% CI 43.0–58.0) and 30% (95% CI 22.0–38.0) for female and male incarcerated individuals, respectively (Fazel, Yoon, et al., 2017).

As described above, a large number of reviews have examined the prevalence rates of mental disorders amongst adults in prison, but fewer have focused on adolescents in the criminal justice system, and there is evidence to suggest that prevalence differs by age group. A comprehensive review of 25 studies including 16,750 incarcerated adolescents by Fazel, Doll, et al. (2008) found that the prevalence amongst these adolescents was higher than that in the general adolescent population, consistent with trends in adults (Baranyi et al., 2019; Fazel, Hayes, et al., 2016). Yet, the authors reported that the prevalence rates of specific psychiatric disorders differed between adolescents and adults in justice settings. Notably, female adolescents had roughly a two-fold greater prevalence of major depression than adult women in prison (29% vs. 12%) (Fazel & Danesh, 2002). However, this review is now out-of-date, since its search for primary investigations concluded in late 2005, and PTSD was not included in the initial search.

Since the review of 2008, many more surveys of psychiatric morbidity in detained adolescents have been published—highlighting the need for updating these estimates to determine if the prevalence of treatable mental disorders has changed over time (Hughes et al., 2020). Such information is necessary for early identification of mental health needs and provision of adequate psychiatric care for justice-involved adolescents, which in turn could generate larger public health benefits, such as youth crime reduction (Fazel, Doll, et al., 2008).

This thesis takes a particular focus on psychiatric disorders amongst justice-involved youth. It is believed that health and developmental difficulties, together with social and economic disadvantages in early childhood and adolescence, have a significant influence on pathways into the criminal justice system, and that incarceration can exacerbate these effects (Fazel, Bromberg, et al., 2022; Hughes et al., 2020). Early intervention and diversion from prisons and other closed environments is essential to mitigate the impact of health, social and economic determinants, and structural drivers of criminal justice involvement, both in adolescence and later in life (Borschmann et al., 2020; Hughes et al., 2020). This has implications beyond the realm of juvenile justice, as adolescence is a time of immense potential for health and the establishment of patterns that will carry into adulthood (S. M. Sawyer et al., 2012).

Unmet health needs and risk of violent reoffending

Incarcerated populations experience elevated burdens of mental disorders, infectious diseases, and premature mortality, mostly attributed to suicide, violence and other non-natural causes of death (Fazel & Baillargeon, 2011). Compared to the extensive literature on mental health and reoffending, few research have studied the effect of physical health on this adverse outcome (Link et al., 2019). Previous studies underscore the importance of physical health for successful reintegration following imprisonment, suggesting that unmet health needs might be associated with recidivism (Link et al., 2019; Schroeder et al., 2011; Stogner et al., 2014). One hypothesis is that poor physical health precludes opportunities for employment and financial stability on release, and places added strain on family relationships (Link et al., 2019). System-level factors have also been understudied compared with individual-level ones, and the evidence thus far is inconclusive, but one nationwide longitudinal study using a dual research design (i.e. between-individual and within-individual) found little differences between prison facilities and the risk of recidivism (Yu et al., 2022).

Although the link between mental illness and violence perpetration is well established, the vast majority of people with mental illness never commit violent crimes (Whiting, Gulati, et al., 2021; Whiting, Lichtenstein, et al., 2021). However, psychiatric and substance use disorders are associated with increased risk of violent reoffending in people released from incarceration. In a Swedish register study, bipolar and substance use disorders were associated with a substantially increased hazard of violent reoffending; assuming causality, the PAF for psychiatric disorders was about 40% in female and 20% in male individuals (Chang et al., 2015). These findings corroborate those from an earlier study (Baillargeon et al., 2009), in which the greatest increase in risk for reincarceration was found amongst persons with bipolar disorder in prison. A study in Italy also found a strong relationship between PTSD symptoms and the risk of recidivism, with the authors suggesting that a trauma-based framework may be useful for identifying risk factors for violent reoffending (Ardino et al., 2013). Considering that people released from prison—especially those with unmet mental health care needs—are at increased risk for victimisation and perpetration of violence, including them as a priority group in violence prevention strategies could contribute to reducing societal violence by counteracting the cycle of violent reoffending (Baillargeon et al., 2009; Chandler et al., 2009; Willoughby et al., 2021).

1.4 Violence risk assessment tools

As was mentioned in the previous sections, the health of people who experience incarceration has clear ramifications for the general public and governmental agencies. Prison health and public health are in many ways intertwined (Dumont et al., 2012); yet, prison health research remains severely underfunded (Ahalt et al., 2015; Kouyoumdjian et al., 2017). Inadequate investment also impacts the accessibility to and the quality of health services received within prisons, which are inferior to that delivered in the community (Wildeman & Wang, 2017). Faced with this reality, prediction model approaches have gained importance to allocate limited resources to individuals at the greatest risk for adverse health and recidivism outcomes.

1.4.1 Clinical prediction models

Clinical prediction models are intended for use by clinicians in patient care and are aimed to facilitate clinical decision-making (Laupacis et al., 1997). These models are designed to make diagnostic (risk that an outcome is currently present) or prognostic predictions (risk that it will occur in the future) related to a given outcome of interest (often a disease or health state) (Steyerberg, 2009; van Smeden et al., 2021). They are developed to address a specific area of clinical uncertainty (Steyerberg et al., 2013), whereby a limited number of patient, treatment or outcome-related predictors (or risk factors) are combined to determine the likelihood of event occurrence within a specified time-frame (Steyerberg, 2009). Their application goes well beyond medicine; in the past few decades, there has been a substantial increase in the number of scientific articles on modelling for prediction (616% increase between 1993 and 2017), especially in applied sciences (e.g. mathematics, engineering and computer science) (Steyerberg, 2009). Whilst several multivariable prediction models are commonly employed in clinical specialties such as cardiology (e.g. QRISK or Framingham, Hippisley-Cox et al., 2017; Hippisley-Cox et al., 2007; Parikh et al., 2008) and oncology (e.g. PREDICT, Thurtle et al., 2019; Wishart et al., 2010), their application is far less widespread in other fields.

There are currently no prediction models in psychiatry that have achieved a similar level of clinical impact to the cardiovascular disease and cancer risk ones. However, as the field increasingly moves toward precision medicine, and electronic health records and large registry data become readily available, prediction modelling approaches in psychiatry are burgeoning (Fusar-Poli et al., 2018). A now outdated qualitative review of risk prediction in mental health (from 2013) yielded 43 relevant studies (Bernardini et al., 2017). By contrast, a recent review of prediction modelling for psychiatric outcomes identified more than 300 prediction models (70% of which were published in the last five years), but found that the current evidence base is limited by methodological issues, including high risk of bias, overfitting and poor reporting (Meehan et al., 2022). Another review of prediction models in psychiatry, with a particular focus on psychosis, highlighted similar barriers to research progress in the field (Salazar de Pablo et al., 2021). Despite this, these models have emerged as a means of providing more empirically grounded and cost-effective psychiatric care (Fusar-Poli et al., 2018), namely related to psychosis (e.g. Fusar-Poli et al., 2017), major depression (e.g. Kessler et al., 2016), suicidality (e.g. Levey et al., 2016) and violent behaviour (e.g. Fazel, Wolf, Larsson, et al., 2017).

Prognostic models can also serve as the framework for structured risk assessment tools (or instruments), some of which are used in forensic psychiatry and criminal justice (M. A. Campbell et al., 2009; Desmarais et al., 2016; Fazel, Burghart, et al., 2022; Fazel et al., 2012; Ramesh et al., 2018; Singh et al., 2014; Singh, Grann, et al., 2011; Singh, Serper, et al., 2011). In 2014, an international survey estimated that more than 200 structured risk assessment tools were regularly employed across 44 countries (Singh et al., 2014). One reason for this interest is that these tools serve many purposes by providing an estimate of future risk for reoffending, including to inform sentencing, parole, release and probation decisions (especially in the US) (Monahan & Skeem, 2016; Van Ginneken, 2019). Recently, Fazel, Burghart, et al. (2022) reviewed criminal risk assessment tools used to support decision-making in justice settings and identified 27 independent validation studies, amongst which were 20 studies (74%) from the US.

Risk stratification also allows for the identification of people at higher risk of reoffending who may be suitable for pharmacological and psychological interventions, and for those at lower risk to forego potentially unnecessary treatment (considering limited resources). As such, tools which incorporate modifiable risk factors, rather than solely unmodifiable (or static) ones, provide therapeutic targets in the reduction of recidivism (Fazel, 2019). Moreover, research has demonstrated that these tools generally provide a more accurate and reliable risk assessment compared to unstructured clinical judgement (Ægisdóttir et al., 2006).

In this thesis, I will focus on the predictive validity of these tools. The implications of the findings for management of people leaving prison will also be discussed. Both beneficial applications of violence risk prediction tools (mainly additional support in prison and on release) and the ethical implications of potentially harmful applications (largely related to sentencing decisions) require consideration (Fazel, 2019). The latter can be mitigated by transparency in the design, development and analysis of these instruments (Jalali et al., 2020). In this line of thought, Salganik (2019) rightfully mentions that "*these new opportunities are also accompanied by new risks*," and this statement will serve as a guiding principle throughout my thesis.

1.4.2 Current limitations in violence risk prediction

A recent controversy concerning the debatable use of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) tool by probation officers and judges to make decisions in regards of parole conditions and bail in the US, highlights the ‘double-edged knife’ potential of these new technologies, because the COMPAS has not released its algorithms, which therefore cannot be validated, critically appraised or understood by those using it. The field of violence risk prediction is limited by methodological shortcomings, lack of transparency in reporting of methods and results, and poor predictive performance (Fazel et al., 2012; Singh, Grann, et al., 2011). These limitations also apply to other mental health applications of prediction models. As such, prediction model development (and internal validation) studies often lack a predefined protocol, an appropriate sample size, adequate missing data methods, and comprehensive information on individual predictors, their weighting, and key model performance measures (Hou et al., 2019; Salazar de Pablo et al., 2021; Senior et al., 2021).

Reporting of most models for violence risk assessment poorly complies with current best practice recommendations and guidelines outlined in the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement and the Prediction model Risk Of Bias ASsessment Tool (PROBAST) (Collins et al., 2015; R. F. Wolff et al., 2019). Typically, discrimination (the ability to distinguish between the presence and the absence of the outcome) is solely described using the area under the curve (AUC or c statistic/c-index), whilst calibration (the agreement between predicted and observed risk estimates) is overlooked, and thereby it is often unclear whether risk predictions are reliable (Cook, 2007; Steyerberg, 2009; Van Calster et al., 2019).

Calibration is highly relevant when evaluating a model’s predictive performance in datasets other than the one used in the original development (referred to as external validation). In practice, an individual’s predicted risk (for a given outcome) may be vastly different from the actual risk if poorly validation models are used (Van Calster et al., 2019). Generalisability is dependent on the quality of the initial prediction model and characteristics of the external validation sample. Specifically, different predictor effects and case-mix (distribution of outcome/predictors between development and validation cohorts) can reduce transportability of existing models in new settings (Steyerberg, 2009). External validation tends to yield more robust and realistic estimates of predictive performance, by limiting overfitting. Yet, very few risk prediction models undergo external validation, and even fewer are ultimately implemented in clinical practice (Siontis et al., 2015).

1.4.3 Oxford Risk of Recidivism (OxRec) tool

In order to address these gaps, researchers from the Forensic Psychiatry and Psychology Research group at the University of Oxford and their international collaborators developed OxRec—a novel, scalable tool for violent reoffending in people released from prison (Fazel, Chang, et al., 2016). This prediction model was derived and validated using national link registries for nearly 50,000 people released from prison in Sweden to predict violent reoffending at one and two years after release. The OxRec model incorporates 14 routinely collected predictors, some of which are modifiable, and their individual weighting has been published. These risk factors can be grouped in three domains: criminal history (length of incarceration, violent index offence, previous violent crime), sociodemographic (sex, age, immigrant status, civil status, education, employment, disposable income, neighbourhood deprivation) and clinical information (alcohol or drug use disorder, any mental disorder, any severe mental disorder) (see Table 1.2 for definitions).

OxRec provides critical information to prison and probation services in making decisions about treatment, release and supervision, and is intended to complement professional judgement, rather than replace it altogether. It aims to facilitate assessment and linkage to care, and in turn assist in treatment allocation of individuals at high risk of violent reoffending, particularly in terms of substance use and mental health community-based interventions, resources for which are often scarce. OxRec is less time-consuming and resource-intensive than other widely adopted tools, such as the HCR-20 (Historical Clinical Risk Management-20; Douglas

Table 1.2: OxRec predictors and their original definitions

Variable	Definition
Sex	Assigned at birth.
Age	Age at release from prison.
Immigrant	First or second generation immigrants (self or either parent born outside of Sweden).
Length of incarceration	Duration of incarceration for most recent offence.
Violent index offence	Most recent offence was homicide, assault, robbery, arson, any sexual offence (rape, sexual coercion, child molestation, indecent exposure, or sexual harassment), illegal threats, or intimidation.
Previous violent crime	Any conviction for a violent offence previous to most recent offence (i.e. before index offence).
Education	Lower secondary, upper secondary, post-secondary.
Employment	Employed at incarceration. (Worked for at least 4 hours [based on their income information] during November before incarceration).
Income	Negative/Zero/Low/Medium/High (Low: 80^{th} percentile). 'Low' and 'Medium' disposable income had accounted for 93%.
Neighbourhood deprivation	Principal components analysis of: mean disposable income, % welfare recipients, % unemployed, % divorced individuals, % with only primary school qualifications, % of immigrants (defined as individuals who were not born in Sweden), residential mobility rate, and crime rate.
Alcohol use disorder	Diagnosis of alcohol use disorder (lifetime: before or during incarceration—ICD-8: 291, 303; ICD-9: 291, 303, 305A; ICD-10: F10).
Drug use disorder	Diagnosis of drug use disorder (lifetime: before or during incarceration—ICD-8: 304; ICD-9: 292, 304, 305 excl. 305A; ICD-10: F11-F19).
Any mental disorder	Diagnosis of any mental disorder excluding substance use disorders (lifetime: before or during incarceration).
Any severe mental disorder	ICD diagnosis of schizophrenia-spectrum or bipolar disorder (lifetime: before or during incarceration).

Note. Table adapted from Fazel, Chang, et al., 2016; Fazel et al., 2019.

and Reeves, 2010), OASys (Offender Assessment System; P. D. Howard and Dixon, 2012), and VRAG/VRAG-R (Violence Risk Appraisal Guide-Revised; Rice et al., 2013). It requires minimal training and can be completed approximately 10 minutes using a freely accessible online calculator: https://oxrisk.com/oxrec-9/. Moreover, the model demonstrated at least comparable levels of predictive performance as the most commonly used structured instruments for violence risk (Fazel et al., 2012), and possibly better performance in external validation (Fazel, Burghart, et al., 2022). The OxRec model has been validated in the Netherlands, reporting adequate predictive accuracy (Fazel et al., 2019), but more external validation research is required to assess its generalisability across other countries. This is particularly needed in LMICs, where the ability of existing tools to predict violent reoffending is unknown, despite most people who experience incarceration living in them.

1.5 Effective interventions for recidivism

1.5.1 Pharmacological treatments

Research on pharmacological treatments of psychiatric disorders has shown some clinical benefits for reducing reoffending risk. Amongst people released from prison in Sweden, a register study using a within-subject design found that the dispensing of psychotropic medications (antipsychotics, psychostimulants, and drugs for addictive disorders) was associated with lower rates of violent reoffending (Chang et al., 2016). A smaller study in England and Wales also showed that pharmacological interventions in prison delayed violent reoffending amongst individuals with current symptoms of psychosis (Igoumenou et al., 2015). Commonly prescribed medications to treat alcohol and opioid use disorder (specifically buprenorphine and methadone) appeared to reduce crime in a population-based study in Sweden (Molero et al., 2018) and another one in Canada (Russolillo et al., 2018); whether these findings translate to persons who have experienced incarceration has not yet been investigated. To date, there is mixed evidence for the effectiveness of opiate substitution treatments in prison for reducing recidivism, with some systematic reviews reporting modest reductions (Hedrich et al., 2012), whilst some others found no beneficial effects (Moore et al., 2019).

1.5.2 Psychological interventions

Several non-pharmacological interventions have also been developed and implemented by prison systems to improve long-term social and health outcomes for people who experience incarceration. Whilst there is some evidence for the effectiveness of common psychological interventions delivered in prison on reducing recidivism (Papalia et al., 2019), the strength of the evidence base is limited by important methodological challenges, particularly when it comes to the quality of primary studies (Koehler et al., 2013). For instance, one of the most comprehensive reviews of prison-based drug and alcohol interventions (both pharmacological and psychological) found that these were effective in reducing recidivism, but that only one psychological intervention study amongst the 49 included studies was methodologically strong (De Andrade et al., 2018).

Few randomised controlled trials (RCTs) have been undertaken, and further research is required to establish whether the evidence holds in larger studies with more rigorous designs (Fazel, Hayes, et al., 2016). As such, an RCT is often regarded as the gold standard for determining the efficacy of therapies since its methods reduce the possibility of bias (e.g. influence of confounding factors) (Akobeng, 2005; Hariton & Locascio, 2018). Moreover, most psychological therapies are tailored to mental health problems and other key modifiable risk factors for recidivism (e.g. housing instability, unemployment, and financial difficulties) are often not considered (Hirschtritt & Binder, 2017). Thus, more work must be done to ensure that psychological treatment in prison is consistent with the evidence (Kinner & Young, 2018), and that justice systems engage in evidence-based policies and practice to reduce violent reoffending.

1.6 Summary and thesis outline

In this introductory chapter, I have provided an overview of violence as a public health concern, the contribution of people who have experienced incarceration to societal violence, various risk factors for interpersonal violence and violent reoffending, and current approaches to violence risk prediction in this population. I argue that more research on the epidemiology of psychiatric disorders in prison, the effectiveness of psychological interventions, and the scientific validity of approaches to structured violence risk assessment is needed. I have also outlined some critical evidence gaps that guide the three main research questions addressed by this thesis:

1. What is the prevalence of mental disorders amongst adolescents in juvenile detention and correctional facilities?

2. What is the predictive validity of a violence risk prediction tool (OxRec) across different geographical settings (Tajikistan and England)?

3. What is the effectiveness of psychological interventions on recidivism after release from incarceration?

Overall, this thesis examines the prevalence, assessment and management of reoffending risk in people in prison. To do so, it draws upon methods from psychiatric epidemiology, public mental health and prediction modelling to investigate risk factors, evaluate the effectiveness of interventions, and validate an existing risk assessment tool in new geographical settings. The overarching aim of the thesis is to provide a body of research that will contribute to improving outcomes for justice-involved individuals and reducing societal violence.

In Chapter 2, I focus on an important modifiable risk factor for violent reoffending in a key population. The chapter provides global estimates on the prevalence of various treatable mental disorders amongst adolescents in juvenile detention and correctional facilities. In this review, I also consider gender differences and sources of heterogeneity between studies.

Chapters 3 and 4 explore the external validity of the OxRec model for assessing risk of recidivism amongst people released from prison. I analyse two data sets that include information on criminal history, sociodemographic and clinical risk factors, to determine if OxRec could be implemented in Tajikistan (LMIC) and England (HIC). This research is the first validate a structured violence risk assessment tool for a criminal justice population in a low- or middle-income country.

In Chapter 5, I focus on the effectiveness of widely implemented psychological interventions delivered in prison to reduce repeat offending on release. This review is the first to solely evaluate the effects of randomised controlled trials. It also improves the precision of overall estimates by comparing effect sizes by study size, control group and intervention type.

In Chapter 6, I provide a general discussion of the empirical findings from this thesis, including their strengths and limitations. Potential implications for clinical practice, service provision and planning, and policymaking in the criminal justice system, as well as recommendations for future research are also addressed.

This chapter is adapted from:

Beaudry, G., Yu, R., Långström, N., & Fazel, S. (2021). An updated systematic review and meta-regression analysis: mental disorders among adolescents in juvenile detention and correctional facilities. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*(1), 46–60.

2

Prevalence of mental disorders in detained adolescents

Contents

2.1 Introduction

In the previous chapter (Chapter 1), I contextualised the research questions underlying this thesis in the light of the relevant literature. I also described the overarching aims and objectives of my research programme. The aim of the present chapter is to update the evidence on psychiatric morbidity amongst adolescents in juvenile detention and correctional facilities and provide current global prevalence estimates of treatable mental disorders in this key population. As described in Chapter 1, prior meta-analytic research has shown that adolescents who experience detention are disproportionately affected by mental illness compared with those in the general population. Available prevalence data have increased substantially over the last decade, and therefore the most recent evidence needs to be synthesised to guide evidence-based practice.

In Western nations, adolescents account for approximately 5% of the prison population, and on any given day in the US, over 48,000 young people are detained in various juvenile facilities (W. Sawyer, 2019). Psychiatric disorders are reported to be widespread among incarcerated youth (Hughes et al., 2020; Schubert et al., 2011). Moreover, it is recognised that psychiatric disorders in this population are associated with an array of adverse outcomes, namely poor prognosis of mental health difficulties, elevated rates of substance misuse (Penn et al., 2003; Teplin et al., 2012), an increased likelihood to experience or perpetrate violence in intimate relationships, and psychosocial difficulties in adulthood. Several studies have also implicated psychiatric disorders in increasing the likelihood of reoffending amongst adolescents (McReynolds et al., 2010; Plattner et al., 2009).

A previous systematic review and meta-analysis synthesised evidence until 2006 on the prevalence of mental disorders in detained adolescents. The findings highlighted considerable unmet mental health needs (Fazel, Doll, et al., 2008). Since then, a significant body of new primary research has been published. However, more recent systematic reviews have been limited by their scope (e.g. by only including English-language reports or not searching the grey literature), a lack of quantitative methods (including heterogeneity analyses), and the use of inconsistent time frames for psychiatric diagnoses (e.g. in past month, past year, and lifetime) (Black et al., 2015; Gottfried & Christopher, 2017). Thus, up-to-date prevalence estimates able to capture the true global burden of mental illness in adolescents who experience incarceration are unavailable.

This chapter presents an updated systematic review and meta-analysis on the prevalence of mental disorders in detained adolescents, thus addressing the first research question of my thesis. Meta-analysis was preferred to alternative methods of synthesis (e.g. narrative synthesis) in this review due to anticipated manageable levels of heterogeneity and the availability of data to derive global prevalence estimates (M. Campbell et al., 2019; Higgins & Green, 2019). I also consider PTSD, which has become increasingly researched in this population in recent years. The findings should inform service provision, planning and future research, and also benefit harm-reduction efforts for recidivism.

2.2 Methods

2.2.1 Protocol and registration

This systematic review was conducted in line with the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) statement (Moher et al., 2009), and the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) guidelines (Stroup et al., 2000). The purpose of these recommendations is to guarantee that research quality and possible sources of bias are considered when selecting observational studies and evaluating the results of meta-analyses. (Brooke et al., 2021). The search protocol was also registered prospectively with the PROSPERO International Prospective Register of Systematic Reviews (CRD42019117111).

2.2.2 Search strategy

I identified studies published between January 1966 and October 2019 reporting the prevalence of mental disorders in adolescents aged between 10 and 19 years in juvenile detention and correctional facilities. For the period January 1966 to May 2006, the methods were described in a previous review conducted by two collaborators (Prof Seena Fazel and Prof Niklas Långström) (Fazel, Doll, et al., 2008). For this update, I searched electronically the following databases: EMBASE, PsycINFO, Medline, U.S. National Criminal Justice Reference System Abstract Database, Global Health and Google Scholar. The search strategy featured terms related to adolescents (juvenile*, adol*, young*, youth*, boy*, or girl*) and custody (prison*, jail*, incarcerat*, custod*, imprison*, or detain*), which was identical to the previous review. For psychotic illnesses, major depression, ADHD and conduct disorder, new search dates ranged from December 2005 to October 2019. However, for PTSD, searches began in January 1980 to coincide with the addition of this disorder to the DSM-III (Gersons & Carlier, 1992). Reference lists were hand-searched. No language restriction was set, and non-English surveys were translated (Figure 2.1).

2.2.3 Study eligibility

Adolescence was defined from the age of 10 to 19 years (UNESCO, 1985), comparable with the previous review and consistent with the *Lancet commission on adolescent health and wellbeing* (Patton et al., 2016). Studies reporting diagnoses of psychotic illnesses, major depression, ADHD, conduct disorder and/or PTSD among adolescents in juvenile detention and correctional facilities were included. Despite being common amongst justice-involved youth, I decided not to include anxiety disorders and substance use disorders, as these two disorders were recently reviewed (Fazel, Yoon, et al., 2017; Livanou et al., 2019). Two other reasons informed this decision; they were not part of the original review (Fazel, Doll, et al., 2008), and the clinical implications of high prevalence of substance use disorders is different than for the severe mental illness in the current review (with varying service models in many jurisdictions).

To be eligible, diagnoses had to be determined by clinical examination or interviews conducted with semi-structured diagnostic instruments (Fazel, Yoon, et al., 2017). Therefore, I excluded surveys that employed exclusively self-report

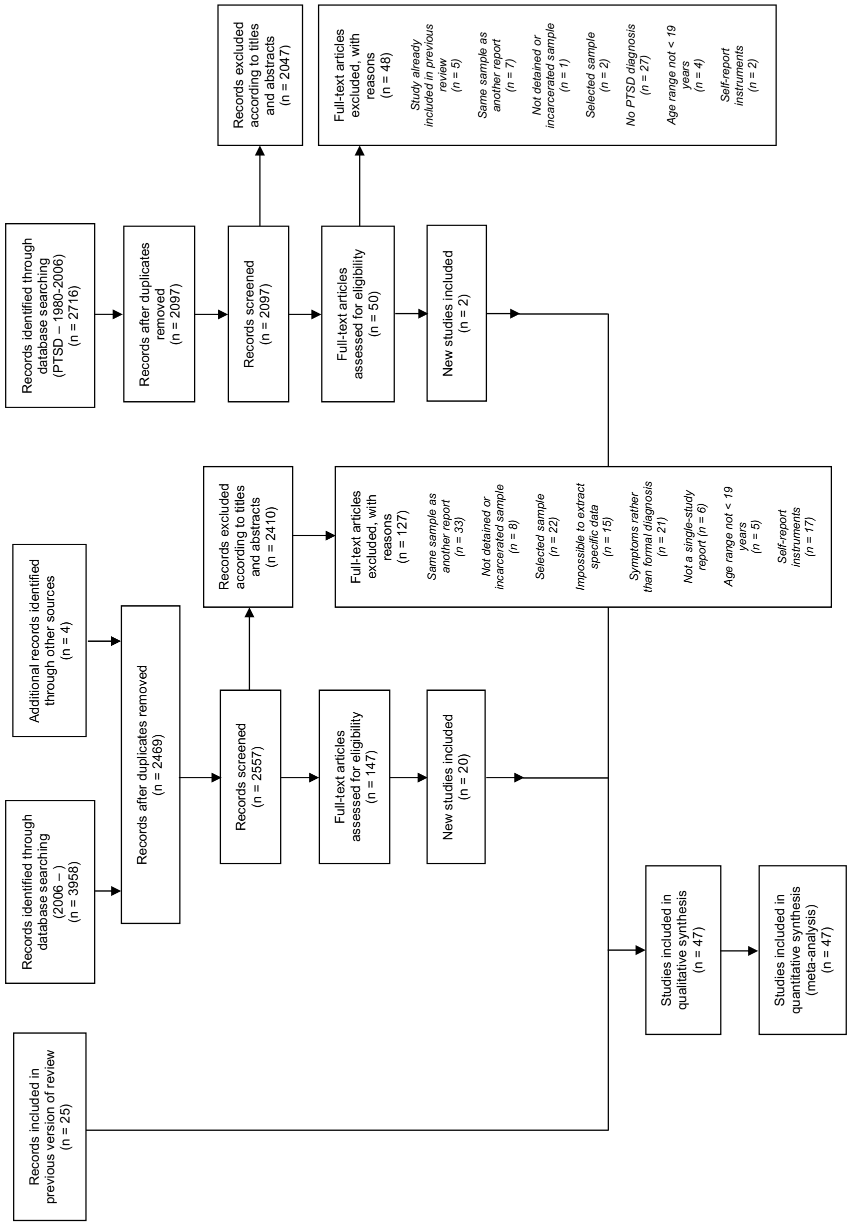


Figure 2.1: Flow diagram detailing the search strategy for the updated systematic review (1966-2019)

instruments to diagnose individuals (but did not include the DISC as it was typically administered in a semi-structured way), in order to mitigate the potential effects of self-report biases on prevalence estimates. It has been suggested that self-reports of mental health issues, and particularly substance misuse, can lead to an erroneous estimation of prevalence (Latkin et al., 2017; O'Malley et al., 1983). Further, previous meta-analyses have shown that laypersons using standardised measures tend to report higher prevalences (e.g. Fazel and Seewald, 2012).

Studies which did not report the prevalence rates of mental disorders separately for male and female adolescents (with the exception of samples including $< 10\%$ of girls) and surveys featuring enriched or selected samples of juveniles in custody were also excluded. Furthermore, included studies reported current prevalence of psychotic illnesses, major depression, ADHD and PTSD, or lifetime prevalence of conduct disorder that adhered to international classifications (ICD and DSM). Thus, one study was partially excluded because the prevalences of psychotic illnesses, major depression and ADHD were reported for the past year rather than the past six months (Colins et al., 2009). Another reason to include PTSD was correspondence from the original review that recommended it to expand the clinical scope (Guchereau et al., 2009). For psychosis, I excluded one small study ($n = 173$) due to being an extreme outlier (11.0%) (Kim et al., 2017).

2.2.4 Data extraction

I extracted data from the newly identified studies according to the fixed protocol used in the previous review and consulted with my academic supervisors in the case of any uncertainty in data extraction. Gender-specific information was collected in regards to prespecified characteristics: geographic location, year of interview, sampling method (consecutive admissions, total population, random, stratified random or some combination thereof), participation rate, number of interviewed adolescents, diagnostic instrument(s) and criteria (ICD or DSM), type of interviewer (psychiatrist versus other), proportion of individuals diagnosed with each disorder, mean age and age range, mean duration of incarceration at the interview, and proportion with violent offences (Fazel, Doll, et al., 2008). Authors of primary studies were contacted when further information was required (Table 2.1).

Table 2.1: Extracted information from included samples, 1966–2019

Study	Country	Population	Type of custody	Sampling strategy	Proportion not consenting	Total number interviewed	Instrument	Diagnostic criteria	Diagnoses reported	Mean age, years	Age range, years	Interviewer	Time detained before interview	Proportion committed violent offences
Bolton, 1976	USA	Juvenile detention center	Not further specified	Stratified random	Not provided	502 males; 149 females	Semi-structured interview	DSM-II	PI	16	16–17	Layperson	4 days	Not provided
Chiles et al., 1980	USA	Juvenile detention center	Correctional	Consecutive (psychotic individuals excluded)	0%	94 males; 26 females	Clinical	Research criteria of depression	MD	Not provided	13–15	Non-psychiatrists	Up to 2 days	Not provided
Kashani et al., 1980	USA	Detention center	Evaluation and detention	Consecutive	Not provided	71 males; 29 females	Clinical	DSM-III	MD	15	11–17	Psychiatrist	Mean 7 days	6%
Hollander and Turner, 1985	USA	Convicted juvenile delinquents	Correctional	Consecutive	8%	185 males	Clinical	DSM-III	PI, ADHD	15	12–18	Staff psychologist and psychiatrist	Not provided	38%
Duclos et al., 1998	USA	Detention center	Not further specified	Consecutive	25%	86 males; 64 females	DISC-2.3	DSM-III-R	MD, ADHD, CD, PTSD	15	12–18	Non-psychiatrists	Not provided	Not provided
Shelton, 1998	USA	Detention facilities	Committal and detention facilities	Complete sample	8%	252 males; 60 females	DISC	DSM-III	PI	16	12–18	Non-psychiatrists	Not provided	Not provided
Ulzen et al., 1998	Canada	Detainees	Secure custodial facilities	Not provided	7%	38 males; 11 females	DICA-R	DSM-III-R	MD, ADHD, CD, PTSD	15	13–17	Research assistant	Not provided	Not provided
Atkins et al., 1999	USA	Central detention facility	Not further specified	Simple random	17%	71 males; 4 females	DISC-2.3	DSM-III-R	ADHD, CD	15	13–17	Social workers, nurses, medical students	Up to 6 months	Not provided
Lader et al., 2000	UK	Detainees	Local prison, Secure juvenile facility (Young Offender's Institution)	Stratified random	2%	314 detainee and 169 sentenced males, and 107 detained/sentenced females	SCAN, Clinical	DSM-IV, ICD-10 (MD)	PI, MD, Mania, BP	Not provided	16–20	Psychiatrists	Modal categories 0–2 months, 6–11 months and 0–2 months	19%

Table 2.1: Extracted information from included samples, 1966–2019 (continued)

Study	Country	Population	Type of custody	Sampling strategy	Proportion not consenting	Total number interviewed	Instrument	Diagnostic criteria	Diagnoses reported	Mean age, years	Age range, years	Interviewer	Time detained before interview	Proportion committed violent offences
Nicol et al., 2000	UK	Detainees	Secure juvenile facility (Young Offender's Institution)	Stratified random	Not provided	51 juveniles (estimate > 90% males)	K-SADS-E	DSM-III-R	PI, MD	Not provided	13–17	Psychiatrist and non-psychiatrists	Not provided	35%
Pliszka et al., 2000	USA	Juvenile detention center	Not further specified	Consecutive	0%	45 males; 5 females	DISC-2.3	DSM-III-R	MD, ADHD, CD, Mania, BP	15	11–17	Non-psychiatrists	Up to 4 days	Not provided
Robertson and Husain, 2001	USA	Detention centers	Secure detention	Simple random	Not provided	168 males; 79 females	APS, JDI	DSM-IV	PI, MD, ADHD, CD, Mania	15	11–18	Mental health workers (non-psychiatrists)	Mean 10.2 days	17% males 18% girls (self-report)
Dimond and Misch, 2002	UK	Remand detainees	Secure juvenile facility (Young Offender's Institution)	Consecutive	5%	19 males	K-SADS-P	DSM-IV	PI, MD, CD, BP	Not provided	15–16	Psychiatrist	Not provided	42%
Gonzalvo, 2002	Spain	Juvenile detention center	Correctional	Consecutive	0%	35 females	Clinical	DSM-IV	PI, MD, ADHD	15	14–17	Psychiatrist	Up to a few days	Not provided
Ruchkin et al., 2002	Russia	Juvenile detention center	Correctional	Complete sample	2%	370 males	K-SADS-PL	DSM-IV	MD, ADHD, CD	16	14–19	Psychiatrist	Not provided	49%
Teplin et al., 2002	USA	Detainees in correctional facilities	Pretrial detention center	Stratified random	4%	1172 males; 657 females	DISC-2.3	DSM-III-R	PI, MD, ADHD, CD, Mania	15	10–18	Trained interviewers (Master's in psychology or associated field)	Up to 2 days	Not provided
Waite D., 2002	USA	Juvenile detention center	Not further specified	Consecutive	0%	9629 males; 1190 females	Clinical	DSM-IV	PI, ADHD, CD	16	11–18	Clinical psychologist	Up to a few days	18% (males) 19% (females)

Table 2.1: Extracted information from included samples, 1966–2019 (continued)

Study	Country	Population	Type of custody	Sampling strategy	Proportion not consenting	Total number interviewed	Instrument	Diagnostic criteria	Diagnoses reported	Mean age, years	Age range, years	Interviewer	Time detained before interview	Proportion committed violent offences
Wasserman et al., 2002	USA	Reception for juvenile delinquents	Assessment before correctional placement	Simple random	3%	292 males	Voice DISC-IV	DSM-IV	MD, ADHD, CD, Mania, PTSD	17	Not provided	Layperson	Mean 18.7 days	36%
Gosden et al., 2003	Denmark	Detainees	Prison and secure social services facility	Consecutive	21%	100 males	SCAN	ICD-10, DSM-IV (ADHD)	PI, MD, ADHD, CD	17	15–17	Psychiatrist	Mean 11 days	86%
Abram et al., 2004	USA	Detainees in correctional facilities	Short-term detention	Stratified random	3%	532 males; 366 females	DISC-IV DSM-IV	PTSD	15	10–18	Trained interviewers (Master's in psychology or associated field)	Up to 2 days	Not provided	
Dixon et al., 2004	Australia	Juvenile detention center	For serious girl offenders	Consecutive	5%	100 females	K-SADS-PL	DSM-IV	PI, MD, ADHD, CD, PTSD	16	13–19	Clinical psychologist	Not provided	71%
Lederman et al., 2004	USA	Juvenile detention	Before trial or long-term placement	Consecutive	27%	493 females	DISC	DSM-IV	MD, ADHD, CD	15	10–17	Non-psychiatrist	Up to 5 days	54%
Vreugdenhil et al., 2004	Netherlands	Six national detention centres	Not further specified	Consecutive	21%	204 males	DISC-IV (DISC-2.3 for PI)	DSM-IV, DSM-III-R (PI)	PI, ADHD, CD	16	12–18	Non-psychiatrists	Mean 4 months	72%
Yoshinaga et al., 2004	Japan	Juvenile Classification Home	Short-term detention	Consecutive	0%	40 males; 8 females	CAPS	DSM-IV	PTSD	17	14–19	Psychiatrist	Up to 4 weeks	Not provided
Abrantes et al., 2005	USA	2 juvenile detention centers	Not further specified	Consecutive	Not provided	218 males; 34 females	PADDI	DSM-IV	PI, MD, CD, Mania, PTSD	16	13–18	Staff (non-psychiatrists)	Not provided	27% (self-report)
Kuo et al., 2005	USA	Juvenile detention center	Secure placement	Consecutive	31%	36 males; 14 females	Voice-DISC	DSM-IV	MD	Not provided	13–17	Non-psychiatrists	Median 4 days	Not provided

Table 2.1: Extracted information from included samples, 1966–2019 (continued)

Study	Country	Population	Type of custody	Sampling strategy	Proportion not consenting	Total number interviewed	Instrument	Diagnostic criteria	Diagnoses reported	Mean age, years	Age range, years	Interviewer	Time detained before interview	Proportion committed violent offences
Chitsabesan et al., 2006	UK	Detainees	Secure juvenile facility (Young Offender's Institution)	Stratified random	7%	118 males; 33 females	SNASA, DSM-IV, PI, MD, ADHD	16	13–18	Psychiatrist	Mean 4 months	Not provided		
Hamerlynck et al., 2007	Netherlands	Detainees	Three Juvenile Justice Institutions	Complete sample	7%	212 females	K-SADS-P-L	DSM-IV	CD	16	12–19	Not provided	Up to 1 month	Not provided
Colins et al., 2009	Belgium	Detainees	Three Youth Detention Centers	Simple random	15%	245 males	DISC-IV	DSM-IV	CD, PTSD	16	12–17	Trained interviewers (Researcher and university students)	Between 3 days and 3 weeks	12%
Indig et al., 2009	Australia	Young people held in custody	Eight Juvenile Detention Centres and one Juvenile Correctional Centre	Simple random	5%	245 males; 39 females	K-SADS-P-L	DSM-IV	PI, MD, ADHD, CD, PTSD	17	13–19	Trained juvenile justice psychologists	Not provided	Not reported for <19 years
Köhler et al., 2009	Germany	Prisoners on remand or in penal detention	Juvenile prison	Complete sample	7%	38 males	SCID (German version)	DSM-IV	PI, MD, CD, PTSD	Not provided	<18	Psychologists	Not provided	75% (Not specific to <19 years)

Table 2.1: Extracted information from included samples, 1966–2019 (continued)

Study	Country	Population	Type of custody	Sampling strategy	Proportion not consenting	Total number interviewed	Instrument	Diagnostic criteria	Diagnoses reported	Mean age, years	Age range, years	Interviewer	Time detained before interview	Proportion committed violent offences
Sørland and Kjelsberg, 2009	Norway	Prisoners	Not further specified	Complete sample	5%	40 males	K-SADS (Norwegian version)	ICD-10	MD, CD	18	15–19	Researcher	60% during 5 first days of custody, 85% during first 18 days, 6 people had been there for longer (from 25 to 240 days)	Not provided
Karnik et al., 2010	USA	Detainees	Dep. of Corrections and Rehabilitation, Division of Juvenile Justice	Consecutive	1%	650 males; 140 females	SCID (PI, MD, PTSD); DICA (ADHD); SIDP-IV (CD)	DSM-IV	PI, MD, ADHD, CD, PTSD	17	<16	Not provided	After 9 months	36%
Gretton and Clift, 2011	Canada	Incarcerated youth	Provincial youth custody centers	Complete sample	Not provided	119 males; 54 females	DISC-IV	DSM-IV	PI, MD, ADHD, CD, PTSD	16	13–18 (females); 12–19 (males)	Trained interviewers with advanced degrees in psychology	Not provided	83% (males); 74% (females)
Mitchell and Shaw, 2011	UK	Remand and sentenced boys	Young Offender Institution	Simple random	7%	115 males	K-SADS	DSM-IV	PI, MD, ADHD, PTSD	17	15–17	Researcher with a significant level of clinical experience	24 hours minimum	53%
Ghanizadeh et al., 2012	Iran	Incarcerated boys	Prison	Not provided	0%	100 males	K-SADS (Farsi version)	DSM-IV	PI, MD, ADHD, CD, PTSD	17	12–19	Researchers	Not provided	83%

Table 2.1: Extracted information from included samples, 1966–2019 (continued)

Study	Country	Population	Type of custody	Sampling strategy	Proportion not consenting	Total number interviewed	Instrument	Diagnostic criteria	Diagnoses reported	Mean age, years	Age range, years	Interviewer	Time detained before interview	Proportion committed violent offences
Harzke et al., 2012	USA	Youth entrants	Youth Commission facilities	Complete sample	Not provided	10469 males; 1134 females	Guided interview structure based on DSM-IV	DSM-IV	PI, MD, ADHD, CD	Not provided	<19	Psychiatrists, doctoral-level clinical psychologists, master's-level associate psychologists in consultation with doctoral-level clinicians, physicians, physician assistants, or nurse practitioners	Up to 30 days	Assault (52.1%), weapons-related offenses (26.8%), robbery (23.4%), threats (11.4%), sexual offenses (6.6%) and murder or manslaughter (3.1%) *Percentages do not add up to 100%
Zhou et al., 2012	China	Detainees	2 Youth Detention Centers	Complete sample	9%	232 males	K-SADS-PL	DSM-IV	MD, DP, ADHD, CD	17	15–17	Psychiatrists	Not provided	73%
Lennox et al., 2013	UK	Adolescent offenders	Young Offender Institution	Consecutive	3%	219 males	K-SADS	DSM-IV	PI, MD, PTSD	17	15–18	Not provided	Between 0 to 26 days	72%
Aida et al., 2014	Malaysia	Detainees	Five prisons that are designated centres for juvenile offenders	Simple random	0%	105 juveniles (estimate >90% males)	MINI-KID	DSM-IV, ICD-10	PI, MD, ADHD, CD	17	14–17	Psychiatrist	Not provided	38%

Table 2.1: Extracted information from included samples, 1966–2019 (continued)

Study	Country	Population	Type of custody	Sampling strategy	Proportion not consenting	Total number interviewed	Instrument	Diagnostic criteria	Diagnoses reported	Mean age, years	Age range, years	Interviewer	Time detained before interview	Proportion committed violent offences
Guebert and Olver, 2014	Canada	Adolescents adjudicated under the Youth Criminal Justice Act or the former Young Offenders Act)	Not further specified	Not provided	Not provided	109 males; 77 females	Diagnostic interview	DSM-IV or IV-TR	MD, ADHD, CD	16	Not provided	Pediatric psychiatrist, registered (usually doctoral level) psychologist	Not provided	83% (males); 74% (females)
Aebi et al., 2015	Austria	Male juvenile detainees	County jail	Consecutive	3%	259 males	MINI-KID	DSM-IV, ICD-10	ADHD, PTSD	17	14–19	Psychiatry residents	up to 4 days	8.5%
Dória et al., 2015	Brazil	Incarcerated boys	Socio-education center	Simple random	Not provided	69 males	K-SADS-PL (Brazilian version)	DSM-IV	MD, ADHD, CD	16	12–16	Trained interviewers	Between 15 to 30 days	Not provided
Lindblad et al., 2015	Russia	Incarcerated delinquents	Juvenile correctional center	Consecutive	2%	370 males	K-SADS-PL	DSM-IV	PI, ADHD, CD, PTSD	16	14–19	Child psychiatrists	Not provided	49%
Aebi et al., 2016	Switzerland	Detainees	Juvenile Detention Center	Consecutive	2%	158 males	MINI-KID	DSM-IV, ICD-10	ADHD, CD, PTSD	17	13–19	Psychiatrists, forensic psychologist	Not provided	63.9%
Kim et al., 2017	South Korea	Juvenile detainees	Male Juvenile Detention Centre	Consecutive	0%	173 males	MINI, K-SADS-PL (Korean version)	DSM-IV, ICD-10	PI, MD, ADHD, CD, PTSD	18	15–19	Clinical psychologist	Not provided	60%
Schorr et al., 2019	Brazil	Juvenile offenders in temporary custody	Provisional Detention Center	Consecutive	0%	74 males	Clinical	DSM-IV	CD	Not provided	15–17	Psychiatrist	Not provided	24% committed homicide offenses

Note. ADHD = attention-deficit/hyperactivity disorder; APS = Adolescent Psychopathology Scale; BP = bipolar disorder; CD = conduct disorder; DICA = Diagnostic Interview for Children and Adolescents (R = Revised); DISC = Diagnostic Interview Schedule for Children; DSM = Diagnostic and Statistical Manual of Mental Disorders (TR = Text Revised); ICD = International Classification of Diseases; JDI = Juvenile Detention Interview; K-SADS = Schedule for Affective Disorders and Schizophrenia for School Aged Children (P = Present, L = Lifetime, E = Epidemiologic); MD = major depression; MINI = Mini-International Neuropsychiatric Interview (KID = for Children and Adolescents); PADDI = Practical Adolescent Dual Diagnostic Interview; PI = psychotic illnesses; PTSD = posttraumatic stress disorder; SCAN = Schedules for Clinical Assessment in Neuropsychiatry; SCID = Structured Clinical Interview for DSM-IV Axis I, II and Personality; SIDP = Structured Interview for DSM-IV Personality; SNASA = Salford Needs Assessment Schedule for Adolescents.

2.2.5 Quality assessment

Study quality was assessed in the included surveys using a modified version of the Newcastle-Ottawa Scale, which appraises for sample representativeness and size, participation rate, statistical quality, and ascertainment of diagnosis (Mata et al., 2015; Stang, 2010). I selected this version of the scale as it was specifically adapted for systematic reviews of prevalence in the context of investigating PTSD amongst people in prison (Baranyi et al., 2018). The potential total score ranged from 0 to 6 points. Studies with a score of 0 to 2 points was considered low quality, 3 to 4 points was considered medium, and 5 to 6 points high quality (Appendix A.1).

2.2.6 Statistical analysis

Statistical model for meta-analysis

Typically, meta-analyses are used to evaluate the effect of interventions; however, they can also be undertaken to determine disease frequency estimates, expressed in terms of incidence or prevalence. Prevalence is understood as the number of cases of a disease in a population during a specific time frame, divided by the population number (Barendregt et al., 2013). Common statistical models for meta-analysis are the fixed-effect and the random-effects models. The former holds the assumption that one true effect size underlies all included studies, meaning that these stem from a single, homogeneous population (Harrer et al., 2021; Nikolakopoulou et al., 2014). Thereby, any differences in observed effects are expected to result from sampling error (Harrer et al., 2021). Conversely, a distribution of true effect sizes is expected in the latter, and thus both within-study estimation error variance and between-studies variance are considered (Borenstein et al., 2010; Higgins & Green, 2019). In this review, random-effects meta-analysis was conducted using the DerSimonian and Laird method to calculate pooled prevalence of each mental disorder, given that heterogeneity was high amongst studies (DerSimonian & Laird, 1986). This method was preferred to others as it does necessitate the assumption of normality for the random effect (Jackson et al., 2010).

A drawback of the random-effects model is that it tends to assign similar weights to all studies, regardless of their sample size (Borenstein et al., 2010). To account for this issue, I aggregated smaller studies, for which the sample size was fewer than 100 individuals. For these small studies, prevalences reported in the text were from the nonaggregated data, whereas the figures were generated using results from the aggregated data. The Poisson distribution was used to obtain 95% confidence intervals as the main outcomes (prevalences of mental disorders) were rare (Lilienfeld et al., 1994). Two studies (Mitchell & Shaw, 2011; Robertson & Husain, 2001), for which the prevalence of psychotic illnesses was zero were imputed according to standard methods (i.e. confidence intervals were calculated using "3" as the numerator and the real population size as the denominator) (Paoli et al., 2002).

Between-study heterogeneity

I reported the Cochran Q and I^2 statistic to indicate the proportion of total variability due to between-study heterogeneity. Cochran's Q represents the weighted sum of squares, whereby residuals are squared, weighted, and then added together to determine the extent to which individual effects differ from the summary effect. The I^2 is derived from Cochran's Q, and it quantifies the effect size variability that is not attributable to sampling error (Harrer et al., 2021). In line with current guidelines, heterogeneity was considered to be low when I^2 ranges from 0 to 40%; moderate from 30% to 60%; substantial from 50% to 90%; and considerable from 75% to 100% (Higgins & Green, 2019). I selected these two measures to assess and report between-study heterogeneity for their complementary benefits; whilst the Q is influenced by the number of studies, the I^2 is less sensitive to changes in statistical power. Moreover, these heterogeneity statistics were preferred to other measures, including the H^2, τ or τ^2, for their increased interpretability (Harrer et al., 2021; Higgins & Green, 2019).

Influence analysis was utilised to identify any influential studies, which would exert a disproportionate influence on overall estimates. I used the Leave-One-Out method, by which pooled estimates of the meta-analysis are recomputed κ times, each time omitting one study (Borenstein et al., 2010; Harrer et al., 2021).

Subgroup analyses and meta-regression

I also conducted subgroup analyses and meta-regression to explore source of heterogeneity on a range of study characteristics (year of publication [$\leq$ 2006 versus > 2006], gender [male versus female], mean age [both as a continuous and dichotomous variable; $\leq$ 15 or > 15 years], sample size [both as a continuous and dichotomous variable; $\leq$ 250 versus > 250 adolescents], study origin [United States versus elsewhere], instrument [Diagnostic Interview Schedule for Children (DISC) versus another instrument], diagnostic criteria [International Classification of Diseases (ICD) versus Diagnostic and Statistical Manual of Mental Disorders (DSM)], interviewer [psychiatrist versus non-psychiatrist], sampling strategy [stratified/non stratified random versus consecutive/complete] and study quality score [both as a continuous and dichotomous variable; high quality studies versus low and medium quality studies]). I first conducted univariate meta-regression, followed by multivariable analysis including factors that reached statistical significance (set at $p < .05$) in the univariate models. To test group differences, subgroup analyses were conducted on all dichotomous variables.

Sensitivity analysis

Approaches to investigate potential publication bias were two-fold. First, funnel plots were generated for visual examination of asymmetry, according to which the effect sizes of interest (prevalence, in this case) were plotted against their standard errors (Lin & Chu, 2018). In the absence of reporting bias, the precision of estimated effects is expected to increase with study size, and the funnel plot should be symmetrically shaped and inverted. Conversely, funnel plot asymmetry occurs in the presence of small-study effects, which can result from a variety of causes, including publication bias (Higgins & Green, 2019). Small-study effects relate to the likelihood that small studies will report greater beneficial effects than large studies (Sterne & Egger, 2001). I also performed Egger's regression test to quantitatively assess for significant small-study effects, which would thereby potentially suggest the presence of publication bias (Egger et al., 1997). Peter's regression test would not have been appropriate in this context, as it is specifically aimed at binary effect size data (Peters et al., 2006).

There are also additional methods of evaluating and controlling for publication bias including the Duval & Tweedie trim-and-fill procedure, PET-PEESE, limit meta-analysis, *p*-curve and selection models (Harrer et al., 2021). However, no technique to date has shown consistent performance benefits when compared to others (Carter et al., 2019; Harrer et al., 2021). Thus, current best practice involves the combination of multiple methods for publication bias (Harrer et al., 2021), which is the approach that I adopted in this meta-analysis. More broadly, I also aimed to counteract potential publication bias in the earlier stages of the review process by actively searching for grey and unpublished literature, such as preprints, dissertations and governmental reports. Such strategy is more effective in accounting for publication bias than any of the aforementioned statistical methods (Harrer et al., 2021). All analyses were done using Stata statistical software, version 17 using metan and metareg commands (StataCorp, 2017). A p value of less than .05 was considered statistically significant.

2.3 Results

I identified 47 studies (46 different samples) from 19 different countries. Through the updated search, I found 22 new surveys. Two studies (Abram et al., 2004; Teplin et al., 2002), were based on the same sample, which provided data for different outcomes. The 47 studies included a total of 32,787 adolescents (28,033 male and 4,754 female adolescents [15%]) of whom the mean age was 16 years (range 10–19 years). Eighteen studies were from the United States (n = 28,018, [86%]), six from the United Kingdom (n = 1,145), three from Canada (n = 408), two each from Australia (n = 384), Brazil (n = 143), Russia (n = 740), and the Netherlands (n = 416), and one each from Austria (n = 259), Belgium (n = 245), China (n = 232), Denmark (n = 100), Germany (n = 38), Iran (n = 100), Japan

(n = 48), Malaysia (n = 105), Norway (n = 40), South Korea (n = 173), Spain (n = 35), and Switzerland (n = 158). These surveys were conducted using a range of sampling strategies: consecutive recruitment of participants (n = 14,768), stratified-random sampling (n = 3,272), simple random sampling (n = 1,432), and complete sampling (n = 12,980). Three studies (n = 335) did not report on their sampling method. Response rates were reported in 38 studies, and only seven of them (n = 1,317) were less than or equal to 75%. Interviews were conducted using the following instruments: 12 used the Diagnostic Interview Schedule for Children and Adolescents, and 14 used the Schedule for Affective Disorders for School-Age Children, Present, Lifetime or Epidemiologic Version, whilst the other surveys employed the Diagnostic Interview for Children and Adolescents, the Research Diagnostic Criteria for Depression, the Adolescent Psychopathology Scale and Juvenile Detention Interview, the Practical Adolescent Dual Diagnostic Interview, the Salford Needs Assessment Schedule for Adolescents, the Mini-International Neuropsychiatric Interview for Children and Adolescents, the Structured Clinical Interview for DSM-IV Axis I, II and Personality, the Clinician-Administered PTSD Scale from DSM-IV, or a semi-structured interview. Most reported diagnoses were assigned using DSM criteria. However, one study provided ICD-10 diagnoses, whilst others combined both DSM and ICD-10 diagnoses. The diagnostic interviews were mostly conducted by psychiatrists, clinical psychologists, researchers and research assistants, or teams with diverse backgrounds. Most studies reported the types of offences and in accordance with previous research (Fazel, Hayes, et al., 2016), I calculated the proportion of adolescents who committed violent offences, which ranged from 6.0% to 86.0%. These study characteristics are also detailed in Table 2.1.

2.3.1 Psychotic illnesses

Twenty-one studies, comprising 27,801 adolescents, reported prevalence of psychotic illness. Overall, 683 of 24,261 male adolescents were diagnosed with a current psychotic disorder (random-effects pooled prevalence 2.7%; 95% CI 2.0%–3.4% [Figure 2.2]). There was substantial heterogeneity between surveys ($\chi^2{}_{17}$ = 71,

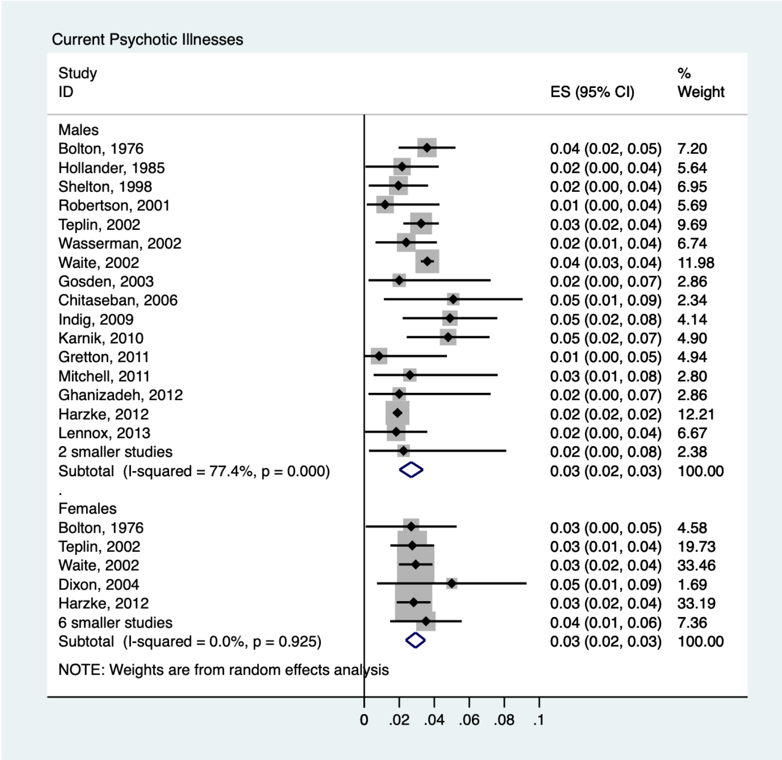


Figure 2.2: Prevalence of current psychotic illnesses amongst incarcerated male and female adolescents. Sample weights are from random-effects analyses. Error bars represent prevalence 95% Confidence Intervals (CI). Smaller studies ($n < 100$) were aggregated.

$p < .001$; $I^2 = 76\%$). Amongst the female adolescents, 105 of 3540 individuals were diagnosed with a current psychotic disorder (random-effects pooled prevalence 2.9%; 95% CI 2.4%–3.5%). Heterogeneity between studies was low ($\chi^2{}_{10} = 5$, $p = .916$; $I^2 = 0\%$). I found no associations between study characteristics and prevalence estimates in meta-regression.

2.3.2 Major depression

I identified 33 studies on major depression in 18,861 adolescents. Overall, 1753 of 15,881 male adolescents (random-effects pooled prevalence 10.1%; 95% CI 8.1%–12.2% [Figure 2.3]) and 774 of 2,980 female adolescents (25.8%; 95% CI 20.3%–31.3%) had a current major depression. There was considerable heterogeneity amongst both male samples ($\chi^2{}_{29} = 339$, $p < .001$; $I^2 = 91\%$) and female samples ($\chi^2{}_{17} = 159$, $p < .001$; $I^2 = 89\%$). meta-regression suggested that both gender and study quality were associated with heterogeneity amongst studies. Male adolescents

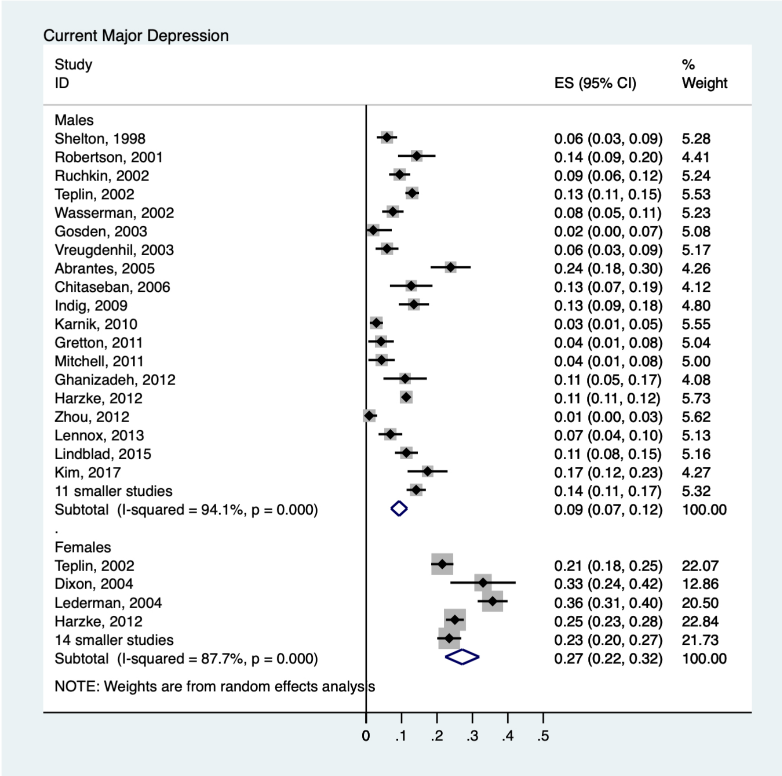


Figure 2.3: Prevalence of current major depression amongst incarcerated male and female adolescents. Sample weights are from random-effects analyses. Error bars represent prevalence 95% Confidence Intervals (CI). Smaller studies (n < 100) were aggregated.

(β = -.14, SE = .032; $p < .001$) and studies with higher quality scores (β = -.08, SE = .036; p = .040) reporter lower prevalence.

2.3.3 ADHD

I identified 27 papers reporting on ADHD amongst 28,749 juveniles in custody. Overall, 4,951 of 24,824 male adolescents (random-effects pooled prevalence 17.3%; 95% CI 13.9%–20.7% [Figure 2.4]) and 836 of 3,925 female adolescents were diagnosed with current ADHD (17.5%; 95% CI 12.1%–22.9%). Heterogeneity was high for male (${\chi^2}_{23} = 824$, $p < .001$; $I^2 = 97\%$) and female samples (${\chi^2}_{12} = 179$, $p < .001$; $I^2 = 93\%$). meta-regression suggested that heterogeneity was partly explained by the publication year (studies published after 2006 reporting a higher prevalence: $\beta = .08$, SE = .04; $p = .03$). In subgroup analyses, the pooled estimate of prevalence of studies published after 2006 was 20.4% (95% CI 17.4%–23.3%) compared to 13.6% (95% CI 8.4%–18.7%) before 2006.

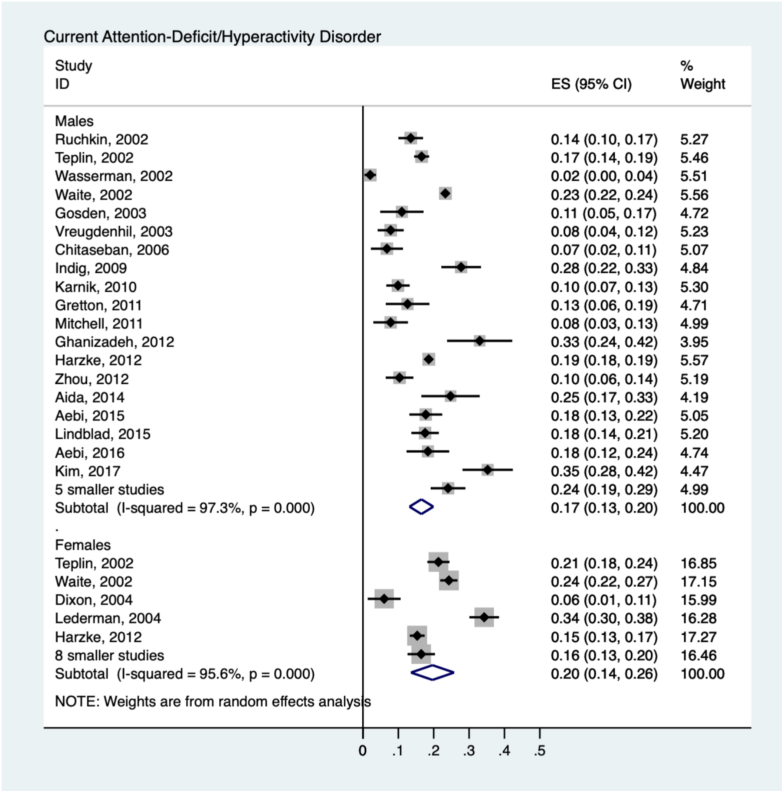


Figure 2.4: Prevalence of current attention-deficit/hyperactivity disorder (ADHD) amongst incarcerated male and female adolescents. Sample weights are from random-effects analyses. Error bars represent prevalence 95% Confidence Intervals (CI). Smaller studies ($n < 100$) were aggregated.

2.3.4 Conduct disorder

I identified 31 studies on conduct disorder in 28,846 juveniles. Overall, 18,042 of 25,184 male adolescents (random-effects pooled prevalence 61.7%; 95% CI 55.4%–67.9% [Figure 2.5]) and 2,226 of 3,662 female adolescents (59.0%; 95% CI 44.9%–73.1%) had a diagnosis of any lifetime conduct disorder. Considerable heterogeneity was observed in male (${\chi^2}_{28} = 2664$, $p < .001$; $I^2 = 99\%$) and female samples (${\chi^2}_{12} = 1127$, $p < .001$; $I^2 = 99\%$).

In meta-regression, studies published after 2006 ($\beta = .19$, SE = .07; $p = .006$) and those with older participants ($\beta = .12$, SE = .05; $p = .013$) had higher prevalences. I also found lower prevalences of conduct disorder where the DISC was used ($\beta = -.22$, SE = .07; $p = .004$). None of these variables remained significant in multivariable meta-regression.

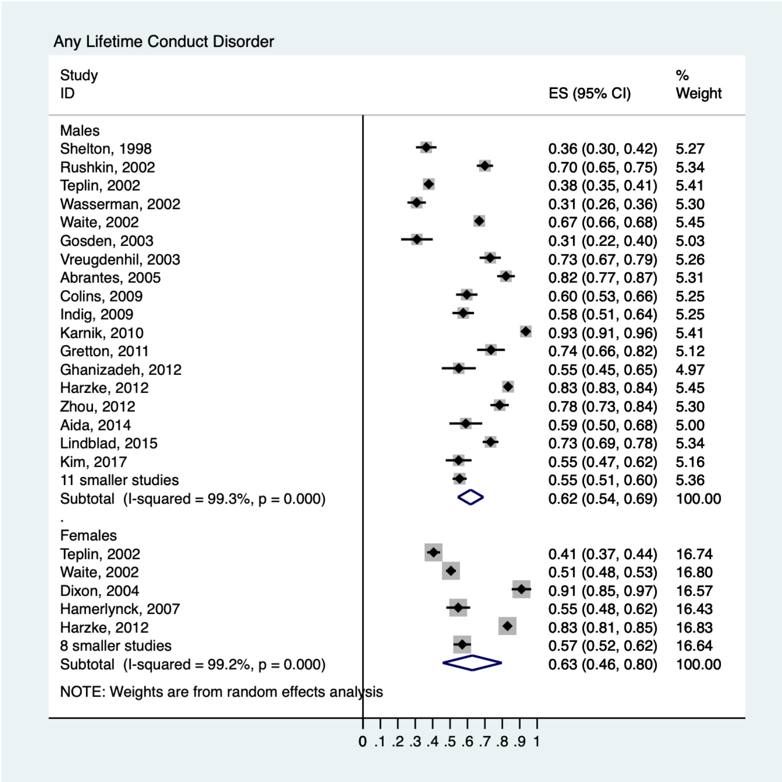


Figure 2.5: Prevalence of any lifetime conduct disorder amongst incarcerated male and female adolescents. Sample weights are from random-effects analyses. Error bars represent prevalence 95% Confidence Intervals (CI). Smaller studies (n < 100) were aggregated.

2.3.5 PTSD

Twenty-one studies reported on PTSD in detained adolescents. Amongst 14,260 male adolescents, 832 (random-effects pooled prevalence 8.6%; 95% CI 6.4%–10.7% [Figure 2.6]) and 334 of 1,876 female adolescents (18.2%; 95% CI 13.1%–23.2%) were diagnosed with current PTSD with heterogeneity in male ($\chi^2{}_{19} = 250$, $p < .001$; $I^2 = 92\%$) and female samples ($\chi^2{}_9 = 41$, $p < .001$; $I^2 = 78\%$). Gender was the only factor associated with heterogeneity in meta-regression (male adolescents had a lower prevalence: β = -.10, SE = .04; p = .01).

2.3.6 Heterogeneity analyses

Figures 2.2–2.6 present gender-specific prevalence estimates. Table 2.2 presents the results from the meta-regression analyses assessing sample characteristics as possible sources of heterogeneity between studies. Influence analysis, which was

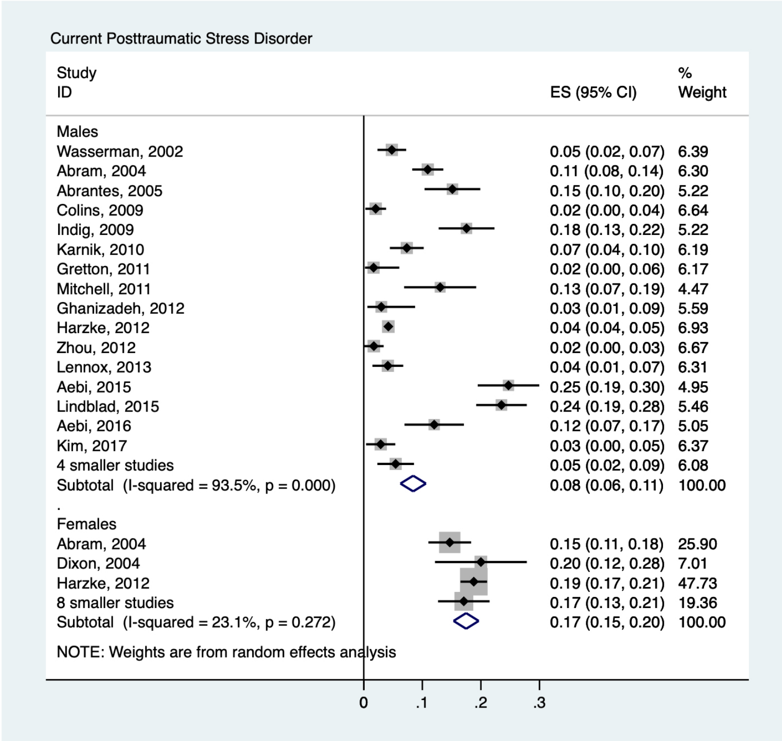


Figure 2.6: Prevalence of current posttraumatic stress disorder (PTSD) among incarcerated male and female adolescents. Sample weights are from random-effects analyses. Error bars represent prevalence 95% Confidence Intervals (CI). Smaller studies (n < 100) were aggregated.

performed by omitting one study at a time, revealed no effect. Egger's regression test showed publication bias in surveys reporting prevalence of conduct disorder (t = -4.98, p = .03) and PTSD (t = 2.32, p = .02), both in male adolescents (see Appendix A; Figures A.1–A.10).

2.4 Discussion

In this chapter, I compiled and synthesised the existing evidence on the prevalence of mental disorders amongst adolescents in juvenile detention and correctional facilities. Forty-seven studies with 32,787 adolescents from 19 different countries were identified in this updated systematic review. I doubled the number of primary studies compared to a 2008 systematic review (Fazel, Doll, et al., 2008). Moreover, I broadened the scope of search from the previous review by adding a new psychiatric diagnosis (PTSD) and more carefully analysed heterogeneity. The prevalence

Table 2.2: Univariate meta-regression analyses examining possible sources of between-study heterogeneity amongst adolescents in juvenile detention

Variable	Psychotic illnesses			Major depression			ADHD			Conduct disorder			PTSD		
	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p
Year of publication: $\leq$ 2006 vs. $>$ 2006	-.005	.004	.22	-.072	.037	.06	.081	.035	**.028**	.194	.066	**.005**	-.029	.039	.47
Gender: male vs. female	-.004	.005	.42	-.144	.032	**<.001**	.002	.040	.96	.028	.079	.72	-.102	.037	**.01**
Mean age (continuous)	-.003	.004	.53	-.033	.024	.18	.003	.027	.91	.124	.047	.01	-.014	.027	.60
Mean age: $\leq$ 15 vs. $>$ 15 years	-.005	.007	.46	-.048	.073	.52	-.022	.079	.79	.182	.163	.27	-.007	.050	.89
Study size (continuous)	.000	.000	.97	.000	.000	.69	.000	.000	.65	.000	.000	.38	.000	.000	.44
Study size: $\leq$ 250 vs. $>$ 250 adolescents	.005	.005	.26	-.022	.045	.63	.002	.040	.96	-.001	.082	.99	.031	.038	.43
Study origin: US vs. elsewhere	.003	.005	.52	.044	.037	.25	-.029	.039	.46	-.094	.073	.21	-.016	.038	.67
Instrument: DISC vs. another	-.005	.005	.33	-.051	.040	.21	-.057	.041	.17	-.218 .071	**.004**	-.071	.038	.07	
Diagnostic criteria: ICD vs. DSM	.006	.005	.20	.034	.074	.64	.008	.080	.92	-.123	.122	.32	-.050	.053	.36
Interviewer: psychiatrist vs. non-psychiatrist	-.006	.005	.19	-.050	.042	.25	-.012	.041	.78	.118	.073	.11	-.004	.045	.93
Sampling strategy: stratified/non stratified vs. consecutive/complete	-.003	.005	.53	-.021	.040	.60	-.010	.042	.81	.099	.080	.22	-.030	.039	.45
Study quality (continuous)	.003	.002	.17	-.029	.013	**.04**	.007	.018	.71	.048	.033	.16	-.004	.017	.81
Study quality: high quality studies vs. low and medium quality studies	.007	.004	.12	-.756	.036	**.04**	-.013	.041	.76	.044	.073	.55	-.003	.039	.93

Note. ADHD = attention-deficit/hyperactivity disorder; DISC = Diagnostic Interview Schedule for Children; DSM = Diagnostic and Statistical Manual of Mental Disorders; ICD = International Classification of Diseases PTSD = posttraumatic stress disorder, SE = standard error. Bold figures are p values < 0.05.

estimates confirm high levels of mental disorders in detained adolescents. The two commonest treatable disorders in male adolescents were depression (about 1 in 10) and ADHD (about 1 in 5). In the female adolescents, approximately 1 in 4 had depression, and 1 in 5 had PTSD. I found higher prevalences of depression and PTSD in girls in custody compared with boys.

2.4.1 Main findings

The review suggests that mental disorders are substantially more common among detained adolescents in comparison to general population counterparts. Approximately 3% of detained adolescents were diagnosed with a current psychotic illness, a tenfold increase compared to age-equivalent individuals in the general population (Costello et al., 2005; Kirkbride et al., 2006). Higher rates of current major depression were found in both male (10%) and female adolescents (26%) in comparison to the general adolescent population (5% and 11% respectively) (Avenevoli et al., 2015). About one out of five detained adolescents had ADHD, compared to one out of ten adolescents in the general population (Visser et al., 2014). Nearly two thirds of detained adolescents were diagnosed with any lifetime conduct disorder whereas the estimated lifetime rate of conduct disorder in US adolescents is approximately 10% (Nock et al., 2006). In addition, adolescents in detention also had higher rates of PTSD than those in the general population, 9% versus 2% in male adolescents and 18% versus 8% in female adolescents (Merikangas et al., 2010). These differences underscore the large burden of psychiatric morbidity in detained adolescents.

Apart from higher prevalence than the general population, the prevalence estimates of mental disorders in adolescent juvenile detention and correctional facilities were also different from those found in adult prison populations. Psychotic illnesses and major depression appear to be more prevalent in adult prison populations than in adolescents custodial populations (Fazel & Seewald, 2012). The prevalence estimates for PTSD are similar in both groups (Baranyi et al., 2018). These comparisons imply that the mental health needs of detained adolescents could be different from those of adult prisoners and may require separate and specifically targeted programs to meet these needs.

The prevalences for ADHD and conduct disorder are higher than in the initial 2008 review. Regarding ADHD, this upward trend may be specific to detained adolescents, as ADHD diagnoses in general population youth have not increased when standardized diagnostic methods are used (Polanczyk et al., 2014). There are two possible explanations for this finding. First, increased prevalence could result from increased awareness of ADHD symptoms amongst health professionals working in custodial services. That is, the true prevalence of these disorders remains unchanged, but clinicians might be identifying them more accurately. Second, higher prevalence may result from improved identification of high risk adolescents over time. Some individuals with ADHD and conduct disorders, whom previously might not be identified are more likely to be selected for placement in custodial correctional facilities due to improved identification of these disorders over time.

Another main finding was the higher prevalence of major depression and PTSD in female detained adolescents compared to their male counterparts. These results are consistent with those from adult prison samples (Baranyi et al., 2018; Fazel & Seewald, 2012; Komarovskaya et al., 2011), and also the general population, military personnel, and terror attack survivors (Cyranowski et al., 2000; Luxton et al., 2010; Parker & Brotchie, 2010). However, the explanations for this specific to incarcerated youth are not clear. Female juvenile delinquency may be more strongly associated with "internalizing" mental disorders than in male adolescents, or girls might be more vulnerable to adverse and traumatic experiences related to an antisocial lifestyle either within or outside the detention centers.

Finally, the funnel plot results suggest publication bias in male adolescents; towards lower prevalence for conduct disorder and towards higher prevalence for PTSD. This could be due to the increased attention that trauma theory has received as a putative causal mechanism for juvenile criminality. In contrast, a highly prevalent descriptive diagnosis such as conduct disorder might be perceived as less useful for etiologic understanding, treatment planning and primary prevention regarding juvenile delinquency.

2.4.2 Implications

Based on the findings of this updated review, there is no pressing need for conducting more primary prevalence studies, especially in high-income countries, considering that the evidence base is quite large and with most prevalence estimates remaining stable over time. However, these disproportionate figures reinforce the importance of systematic mental health screening and linkage to effective pharmacological and psychological treatments upon arrival to custody. Hence, future research could move to treatment and interventions in custodial settings, and investigate modifiable risk factors for adverse outcomes within custody such as self-harm and violence that may be associated with mental health problems. To date, the evidence on psychiatric disorders and violence amongst youth released from detention is limited to a single study (Elkington et al., 2015), which showed that only SUD predicted subsequent violence, thus suggesting an age-dependent relationship between psychiatric disorders and violence (Whiting, Gulati, et al., 2021; Whiting, Lichtenstein, et al., 2021). Effective treatment will likely improve prognosis and reduce suicidality, violence, and reoffending risk (Wibbelink et al., 2017).

2.4.3 Strengths and limitations

The strengths of this chapter include: overall adherence to PRISMA/MOOSE guidance, prespecified, transparent and reproducible methods (see PROSPERO protocol), a comprehensive search strategy, and detailed assessment of the methodological quality of individual studies, between-study heterogeneity (using subgroup analyses and meta-regression), and publication bias. Nonetheless, some limitations should be noted. First, due to discrepancies in how substance use disorder and other mental disorders were classified between studies, it was not possible to reliably examine psychiatric comorbidity. As adolescents who suffer from comorbid disorders generally present an elevated criminogenic risk, future research on comorbidity is needed (Colins et al., 2011; Lindblad et al., 2015; Teplin et al., 2002).

Second, there were no sufficient data on the type of facilities (pretrial versus post-sentencing; short-term versus long-term) where youth were detained. Therefore, I could not explore whether this variable was associated with heterogeneity. Future studies should consider reporting this information on juvenile justice facilities. Third, my analyses were solely based on formal diagnoses of mental disorders, according to the DSM and ICD which provide standard ways of communication between mental health professionals. However, I did not report on subthreshold psychiatric symptomatology, which future work could examine, as these individuals could benefit from preventative programs. An additional limitation from this review is that, the quality appraisal scale was not specifically designed for the purpose of prison prevalence studies, and therefore some of the scoring made assumptions that need further examination (including a lower score for interviews conducted by lay persons using standardized measures versus unstructured clinical interviews conducted by psychiatrists or psychologists, although most of the latter also used standardized tools).

Further, there were high levels of between-study heterogeneity. This is expected due to the differences in jurisdictions about who they detain, availability and effectiveness of health care services, and prison environments. Therefore, further work could examine prevalence rates longitudinally in the same individuals to study trends over time. Moreover, I primarily used US general population data as a point of comparison for the calculated pooled prevalences due to similar diagnostic instruments, age ranges and prevalence periods (Avenevoli et al., 2015; Merikangas et al., 2010; Nock et al., 2006; Visser et al., 2014). Nevertheless, as worldwide rates differ, including for ADHD between high-income countries, prevalences should be interpreted in relation to national or regional general population prevalences. Finally, it was notable that all included studies were conducted in high- and upper middle-income countries despite the global search. Determining whether new research in other countries is required will need to be balanced by information in this review, local needs, and whether such research can be linked to improved services. Region-specific reviews with search strategies adapted to reflect local terminology could be undertaken to broaden the evidence and obtain prevalence estimates that are representative of LMICs (Atilola, 2021). As such, Atilola et al., 2020 have since conducted a systematic scoping review of the psychiatric morbidity amongst justice-involved adolescents in sub-Saharan Africa using specific search terms (e.g. borstal), in which they identified 26 relevant studies (none of which met the inclusion criteria for this review).

2.5 Conclusion

2.5.1 What this chapter adds

This empirical chapter aimed to address the first research question of my thesis, as outlined in Chapter 1. I systematically reviewed and updated the scientific literature on the prevalence of mental disorders in detained adolescents, and confirmed high rates in this key subgroup of the global prison population. Overall trends in prevalence estimates have remained relatively stable over time (Fazel, Doll, et al., 2008), and results reflect the differential prevalence of treatable mental disorders by gender and population group (adolescents vs. adults who experience incarceration) (Fazel, Hayes, et al., 2016). The findings underscore the importance of equitable access to tailored mental health services and effective treatment for justice-involved youth. Such treatment will likely improve prognosis of this population, almost all of whom will reenter the community, decrease their risk of repeat offending, and reduce the substantial social and financial costs related to imprisonment (Cuellar et al., 2004). Taken in the broader context of this thesis, this review provides clear evidence of high rates of treatable mental disorders in prisons and other closed settings. These are associated with poor outcomes on release, including violent reoffending (see Chapter 1). In the next three chapters of my thesis, I will address how to best improve the assessment of recidivism risk and what treatments can be implemented at scale to reduce this risk.

Contribution statement

This chapter is based on the published work: *Beaudry, G., Yu, R., Långström, N., & Fazel, S. (2021). An updated systematic review and meta-regression analysis: mental disorders among adolescents in juvenile detention and correctional facilities. Journal of the American Academy of Child & Adolescent Psychiatry, 60(1), 46–60.* I conducted the data search, extraction, analyses, and prepared the tables and figures under the supervision of Prof Seena Fazel and Dr Rongqin Yu. I also wrote the first draft of the manuscript and implemented the contribution of the co-authors and external reviewers up to final publication.

I am grateful to the following researchers who kindly agreed to provide additional data from their studies: Marcel Aebi, PhD, of the University Hospital of Psychiatry, Department of Forensic Psychiatry, Zurich (Switzerland), Robert J. W. Clift, PhD, of the Youth Forensic Psychiatric Services and the University of British Columbia, Vancouver (Canada), Devon Indig, PhD, of the School of Public Health and Community Medicine, University of New South Wales, Sydney (Australia), Denis Köhler, PhD, of the University for Applied Sciences Düsseldorf, Düsseldorf (Germany), Mark Olver, PhD, of the University of Saskatchewan, Saskatoon (Canada), and Coby Vreugdenhil, MD, PhD, of the Pieter Baan Centrum, Almere (Netherlands).

2.5.2 Next chapter

The two subsequent chapters revolve around the potential generalisability of an existing violence risk assessment tool (OxRec) to new settings. In Chapters 3 and 4, I will directly address the second research question of this thesis, by investigating whether a violence prediction model first developed in Sweden is applicable to Tajikistan (an LMIC) and England (another HIC) based on predictive validity.

This chapter is adapted from:

Beaudry, G., Yu, R., Alaei, A., Alaei, K., & Fazel, S. (2022). Predicting violent reoffending is individuals released from prison in a lower-middle-income country: a validation of OxRec in Tajikistan. Frontiers in Psychiatry, 13(805141), 1–9.

3

OxRec external validation in Tajikistan

Contents

3.1 Introduction

In previous chapters, I have established that violent reoffending amongst people released from prison is a form of predictable and preventable interpersonal violence. I have discussed various risk factors of this public health issue, as well as its wider repercussions for society (see Chapter 1). Chapter 2 found that mental disorders, which are associated with violent reoffending in this population, are particularly prevalent in detained adolescents compared with general population counterparts of similar age. Findings from this chapter also highlighted the differential prevalence distribution of some mental disorders between the adult and adolescent prison population.

Whilst it is clear that access to mental health care services remains scarce and difficult for people in prison, and that psychological interventions delivered in these settings require improvement, there is also a need for more evidence-based allocation of existing limited resources. As described in Chapter 1, implementing risk stratification approaches may help to optimise resource allocation in custodial settings. Yet, most prediction models that incorporate multiple predictors to identify individuals at risk of violent reoffending present significant limitations. Almost all have been developed without predetermined protocols, do not report a range of recommended performance measures, and rarely include modifiable risk factors (Fazel, 2019; Fazel, Chang, et al., 2016). Moreover, violence risk assessment tools are seldom externally validated in new populations (and countries) to assess their predictive performance (Siontis et al., 2015)—and this is especially true in LMIC settings. Thus, it is not known whether these tools are applicable to LMICs, given potential differences in effects of risk factors and base rates of reoffending in various countries (Yukhnenko et al., 2019).

More broadly, despite the fact that more than two thirds of the global prison population live in LMICs (Fair & Walmsley, 2021), the majority of prison health research has been based on data from HICs. Conducting research in prisons is key to addressing public health challenges in these populations, such as mental health, substance misuse and infectious diseases (Ako et al., 2020)—all of which have direct and indirect effects on violence perpetration. Yet, such efforts are met with many challenges, from funding to operational constraints (Kouyoumdjian et al., 2015), and evidence on people in prison in LMICs is limited (Gureje & Abdulmalik, 2019).

In this chapter, I report an external validation of the OxRec tool using data from a prospective cohort study in Tajikistan. Tajikistan is the poorest country in Central Asia, and one of the few LMICs in this region, with nearly 27% of the population living in poverty (Asian Development Bank, 2021). Such a validation could provide an approach to improve risk stratification and guide limited resource allocation in this LMIC setting.

3.2 Methods

3.2.1 Ethical approval

Ethical approval for this validation study was obtained from the Institutional Review Board (IRB) of the Tajikistan Prison Organization on August 8, 2019 and prior to my involvement in this research project. The study protocol is presented in Appendix B.1.

3.2.2 Study design

This OxRec validation study was a prospective cohort study that included individuals in prison. Enrolment took place at the two main gender-specific prisons of Tajikistan, respectively located in the cities of Dushanbe (for men) and Vahdat (for women). Collaborators based in Tajikistan identified participants during two recruitment phases. The first phase took place from August 2019 to December 2019, and the second from January 2020 to March 2020. To investigate the risk of violent reoffending, participants were followed up from the date of their release until the outcome first occurred or the end of the study (within 12 months).

3.2.3 Participants

All people in prison aged 18 years and older, set to be released within the next six months (regardless of their index crime or the length of their sentence), were eligible for recruitment. Prison staff randomly selected eligible persons, with the aim to recruit approximately 1000 individuals. Following this, eligible people in prison were asked whether they wanted to take part in the study. Participation was voluntary. Oral informed consent was obtained from each study participant in keeping with recommended practice in Tajikistan.

The data were collected by trained research staff using structured interviews and criminal records based on predetermined variables. The questionnaire comprised of 44 questions grouped into three distinct categories (i.e. risk assessment tool [OxRec], General Health Questionnaire [GHQ-12], and Self-Reporting Questionnaire [SRQ-20]). The GHQ-12 was produced by a private entity, and as a result, it is protected by copyright. By contrast, the SRQ-20 was developed by the WHO and thereby no copyright fees are associated with its use (Das-Munshi et al., 2020). The reason for selecting the GHQ-12/SRQ-20 to identify mental disorders was to provide a feasible, time-efficient and cost-effective screening procedure for future implementation in LMICs, where there are often a shortage of specialist mental health professionals (Anjara et al., 2020). The questionnaire was first developed in English, and then translated into Tajik and Russian for the purpose of this study. The accuracy of the translation was verified by an independent bilingual expert.

The questionnaire was the basis for the structured interviews. All interviewers were experienced clinical health care professionals from Tajikistan's Department of Health and Prison Organisation who had not previously worked in the two selected prisons. They received training and supervision throughout the data collection process from the Institute for International Health and Education (IIHE). Participants were interviewed individually in the language of their choice—their native language being either Tajik or Russian—and interviews lasted approximately an hour. Results were recorded manually on paper, and then transferred into an anonymised and protected Excel file. Each participant was attributed a unique identification code, allowing for questionnaire data to be linked with recidivism data at the end of the study period.

3.2.4 Outcome and predictors

I defined the predictor and outcome variables in accordance with the original OxRec study (see Chapter 1). In the derivation study, predictors were modelled based on routinely collected data in national linked registers from Sweden (Fazel, Chang, et al., 2016). Therefore, some operationalisation modifications were required to ensure that the OxRec tool was adapted to the local context (see Appendix B Table B.1). Specifically, employment status was expanded to include Tajik labour migrants considering that labour migration is common in Tajikistan (predominantly to Uzbekistan and Russia). For disposable income, previous categories based on percentile points of the Swedish income distribution (i.e. negative, zero, low [$< 20^{\text{th}}$ percentile], medium [20–80^{th} percentile] and high [$> 80^{\text{th}}$ percentile]) were adjusted to better reflect the structure of LMIC economies. As census data on personal income was unavailable, I used the self-reported monthly gross salary to create two new categories: low (below or equal to the international extreme poverty line for LMICs as specified by the World Bank; 1.90 US dollars per person per day) vs. stable (above this line). There was no readily available data on psychiatric history in the form of ICD or DSM diagnostic codes, and thus I relied on self-reported information on alcohol and drug consumption during the past 12 months as a proxy for diagnosis of substance use disorders. As for other mental disorders (excluding substance use disorders), these predictors were assessed using the GHQ-12 and the SRQ-20. A total score of 21 or more on the GHQ-12 was considered as the appropriate threshold for any mental disorder, whereas a total score of 16 or more on the SRQ-20 was taken as being consistent with any severe mental disorder. These cut-offs were based on the prevalence of these predictors in the development study, as there are currently no reliable prevalence estimates of psychiatric morbidity in Tajikistan. I used two distinct tools to reduce the chance of collinearity amongst predictor variables and to ensure that the predictive properties of the OxRec model were retained.

The primary outcome was violent reoffending within one year of release from prison. I could not evaluate reoffending outcomes at two years due to lack of sufficient follow-up data at the time of conducting analyses. Any crime relating to interpersonal violence (including sexual offences, robbery, illegal threats and intimidation) for which the sentence resulted in imprisonment, was defined as violent. This differed from the model development in that violent reoffending was previously measured using conviction data, but this was not possible here. The reason for this being that I did not have access to individual-level data on all convicted individuals in Tajikistan during the study period, but rather only for those who received jail or prison sentences. Time until reimprisonment was measured from the release date onwards. Outcome data were retrieved from criminal records that include information from both police departments and correctional facilities in Tajikistan (i.e. detention centres ['sizo'] and prisons).

3.2.5 Sample size

I assumed a 10% violent reoffending rate for the sample size calculation in keeping with the risk cut-off used in the original derivation study (Fazel, Chang, et al., 2016). The effective sample size was determined by the rule of thumb of 100 events or nonevents (whichever being less frequent) required to detect substantial differences in prognostic model performance (Collins et al., 2016; Harrell et al., 1996; Steyerberg, 2018; Vergouwe et al., 2005). This benchmark can be used in lieu of formal sample size calculations, as these are not necessary for external validation studies. However, a systematic review from 2014 found that 46% of external validation studies published in core clinical journals had an inadequate sample size (and number of events) (Collins et al., 2014). Thus, in this particular study, the targeted minimal sample size was 1,000 individuals released from prison to ensure that at least 100 recurring violent crime events would be recorded.

3.2.6 Missing data

Missing data are of significant concern in prediction modelling research (and medical scientific research more broadly) and identifying the underlying mechanism of missing values of predictors is crucial for selecting appropriate missing data methods (Steyerberg, 2009). Complete-case analysis—whereby individuals with any missing value on covariates or the outcome are omitted—is not only inefficient from a statistical perspective, in that it leads to a decrease in sample size, but also impacts on analysis accuracy, often resulting in bias. Despite this, Collins et al. (2014) found that this simple method was widely employed amongst validation studies. Some degree of missingness is to be expected when validating a model in a different geographical setting, and this usually requires more advanced imputation methods, such as multiple imputation. Further details on multiple imputation for deriving prediction are described in the next chapter (Chapter 4).

As described above, due to considerable differences between the Tajik and Swedish health care and criminal justice systems and issues pertaining to data availability, some definitions required adaptation (Appendix B Table B.1). Moreover, two predictors (i.e. immigrant status and neighbourhood deprivation) were entirely omitted from this OxRec validation study as a result of being unavailable in the Tajikistan data set. Considering that these are the two weakest predictors, the impact of their omission on the estimated model performance was minimal and reduced by imputation to deal with missingness. Therefore, I averaged out these cohort-wide missing variables using mean imputation (i.e. assigning all participants the average value from the derivation sample). There were no missing data for individual participants on the remaining predictors, as indicated by the first two methods.

3.2.7 Statistical analysis

External validation

There are particular challenges related to the validation of survival models, such as censoring. Censoring occurs when there is incomplete information about an individual's survival time or outcome. I applied administrative censoring (i.e. right-censoring) at one year, and individuals who were censored before one year—as a result of insufficient follow-up duration or being lost to follow-up—were omitted from subsequent model performance analyses. More advanced approaches to dealing with censored observations in Cox models have since been proposed (McLernon et al., 2022), and should be considered for future validation studies.

In line with previous work (Royston & Altman, 2013; Steyerberg, 2009; Su et al., 2018), I employed an incremental approach for validating models for time-to-event data previously detailed in a validation study of OxRec in the Netherlands (Fazel et al., 2019). Baseline characteristics of the validation cohort (Tajikistan) were first compared to the derivation cohort (Sweden), using summary measures (Table 3.1). If there were significant differences in definitions, they were adapted to the validation cohort, to improve usability, and thus future clinical utility in the intended implementation context (i.e. Tajik sizos and prisons). I then calculated the OxRec risk score for each participant using the following Cox proportional hazards regression equation:

$$h(t) = h_0(t)^{exp(\beta_1 * x_1 + \beta_2 * x_2 + ... + \beta_p * x_p)}$$

where $h(t)$ is the expected hazard at time t, $h_0(t)$ is the baseline hazard function (often computed using the mean values of predictors) (Steyerberg, 2009), and β is the regression coefficient corresponding to each predictor, as denoted by x. The weighted sum of the predictor variables in the model, where the weights are the regression coefficients, constitutes the linear predictor (or prognostic index) (Royston & Altman, 2013). The values of the linear predictor (including the contribution of each predictor to the OxRec model) and the baseline hazard function can be found in the Supplementary appendix of the original derivation study (Fazel, Chang, et al., 2016).

For OxRec, the Cox regression model is used to estimate the relative hazard of violent reoffending amongst people released from prison. Whilst this semi-parametric model is dependent on underlying assumptions such as the linearity and additivity of predictors, similar to other statistical models, the most important one relates to proportional hazards. The latter presupposes that the relative hazard holds constant over time albeit varying predictor values (Kuitunen et al., 2021; Royston & Altman, 2013). I tested for this assumption using both analytical and graphical methods, with scaled Schoenfeld residuals and visual inspection of the

Table 3.1: Baseline characteristics of the Tajik sample compared with those of the Swedish sample

Variable	**Tajik sample (n = 970)**	**Swedish sample (n = 37,100)**
Sex		
Male	846 (87%)	93%
Female	124 (13%)	7%
Age	Median 35	Median 36
	IQR 28–43	IQR 27–46
Length of incarceration		
< 6 months	701 (73%)	69%
6–12 months	47 (5%)	16%
12–24 months	64 (7%)	10%
≥ 24 months	158 (16%)	4%
Violent index offence	608 (63%)	38%
Previous violent crime	105 (11%)	53%
Civil status		
Other	578 (60%)	35%
Unmarried	392 (40%)	65%
Education		
< 9 years	57 (6%)	48%
9–11 years	784 (81%)	46%
≥ 12 years	129 (13%)	6%
Employment		
Unemployed	345 (36%)	75%
Employed	625 (64%)	25%
Income		
Low	423 (44%)	Negative — < 1% Zero — 6% Low — 53%
Stable	547 (56%)	Medium — 40% High — 1%
Alcohol use	358 (37%)	22%
Drug use	84 (9%)	23%
Any mental disorder	470 (48%)	22%
Any severe mental disorder	44 (5%)	3%

Note. Data are median (IQR) or n (%). Income was calculated using the international poverty line for low-income countries ($1.90 US per day).

Kaplan-Meier curves (Grambsch & Therneau, 1994; Kaplan & Meier, 1958), and found no clear violation. The Schoenfeld residuals could also have been plotted to test model fit and identify covariate values that deviate from the norm. However, one graphical approach was deemed to be sufficient in the absence of proportional hazards assumption violation (Sestelo, 2017).

3.2.8 Model performance

Various measures were employed to assess the validated model performance in terms of accuracy, discrimination and calibration. Accuracy is a measure of overall performance which considers calibration and discrimination simultaneously. OxRec's performance in the Tajikistan prison sample was quantified using the Brier score, whereby the model predictions are compared with the actual outcome (Steyerberg, 2009). In practice, the Brier score is the average mean squared difference between these values, and can range from 0 (best performance) to 1 (worst performance) (Steyerberg et al., 2010).

Discrimination refers to the model's ability to distinguish individuals with the outcome of interest (i.e. violent offence after release from prison) from those without this outcome (Steyerberg, 2009). In prediction models for survival outcomes, this specifically relates to the capacity to differentiate between survival curves for individuals or groups (Royston & Altman, 2013). To evaluate the model's discrimination, the area under the receiver operating characteristic curve (AUC-ROC) was used (Debray et al., 2015). The AUC takes values between 0.5 and 1, and represents the probability that individuals who commit violent crimes will be given a higher risk score than those who do not violently reoffend (Fazel, 2019). The ROC curve is a graphical representation of the relationship between the true positive rate (sensitivity) and the false positive rate (1 – specificity) across a range of classification thresholds. It provides insight in identifying relevant cut-off points for implementing the tool in clinical practice (Florkowski, 2008).

I also calculated sensitivity, specificity, positive and negative predictive values (PPV and NPV) for various prespecified risk thresholds to inform potential benefits (and harms) for intended management purposes. These measures relate to the tool's ability to identify correctly individuals with (sensitivity) and without (specificity) violent reoffending outcomes. Unlike the latter, PPV (the proportion of individuals labelled as high risk who violently reoffend) and NPV (the proportion of individuals labelled as low risk who do not violently reoffend) vary with the outcome prevalence in a given sample (Fazel, 2019; Steyerberg, 2009).

Conversely, calibration indicates the extent to which predicted and observed outcomes are in agreement, and is often underreported in validation studies (Van Calster et al., 2019). This is exemplified in a review by Fazel, Burghart, et al. (2022) on the predictive performance of sentencing risk tools that found only two of the 36 included studies reported such information. In this validation study, calibration was assessed by graphical inspection of the calibration plot, and calculation of the calibration slope, CITL and the expected/observed (E/O) ratio (Steyerberg, 2009). To generate the calibration curve, the predicted event probabilities were plotted against the observed outcome frequencies for each risk decile. The calibration slope assesses the predicted risk distribution; values of < 1 and > 1 respectively suggest that risk estimates are either too extreme or too moderate (Van Calster et al., 2019). CITL compares the average predicted risk with the outcome incidence, thus indicating whether there is a systematic underestimation/overestimation in predictions (Steyerberg & Vergouwe, 2014). The E/O statistic measures the predicted number of events against the observed number of events in the target population. This metric is useful to evaluate the goodness-of-fit of the model, with a value of 1 indicating perfect calibration (Hanson, 2017). I report all performance metrics for each step of the validation process for transparency and reproducibility purposes.

3.2.9 Model updating

Model updating was commensurate to the predictive performance of OxRec as indicated by the above performance measures. First, I conducted simple validation, by which OxRec was applied to the validation data from Tajikistan, using the baseline risk, regression coefficients and linear predictor values calculated in the derivation data. In the presence of inadequate calibration, two recalibration approaches could be employed—the first being updating only the baseline risk, and the second, updating both the baseline risk and recalibrating the model coefficients via a single multiplicative recalibration value. These two steps do not impact the model's discriminative power, as they do not alter the relative effects of predictors on the outcome (risk of violent reoffending). Finally, individual parameters can be re-estimated in the event of predictor effects in the validation sample which diverge significantly from that of the derivation sample, but this final step should only be performed as a last resource. Specifically, this would involve refitting to the validation data the Cox model (from the derivation data) to recalculate the specific regression coefficients (Royston & Altman, 2013). If re-estimation was necessary for multiple predictors, this could indicate the need for developing an entirely new model, rather than validating the existing one (Steyerberg, 2009). Data analysis was undertaken with Stata version 17 (StataCorp, 2017). Significance was set at a threshold of $p < .05$. I reported the results and findings of this study according to the TRIPOD statement (Collins et al., 2015).

3.3 Results

3.3.1 Participants

In this prospective multicentre study, 6,853 individuals were incarcerated, 2,225 of which were eligible, in two Tajik prisons between August 2019 and March 2020. Approximately half were randomly selected (1,123 people in prison), and 1,003 provided consent for study participation. All participants had complete data on the predictor items from the structured interview. Data on one-year post-release outcomes was missing for 33 individuals, who were thereby censored. Thus, the final validation sample included 970 people released from prison.

Compared with the Swedish derivation sample, the Tajik validation sample showed large differences in most predictors, as indicated in Table 3.1. The Tajik cohort included a higher proportion of female participants (13% in Tajikistan vs.

7% in Sweden), individuals with formal education (at least high school; 94% vs. 52%) and employment (64% vs 25%). There were less unmarried Tajik individuals than Swedish ones (40% vs 65%). Moreover, the distribution of the length of incarceration was inverted, with shorter periods of detention being less prevalent (< 1% in Tajik sample vs. 69% in Swedish sample [for less than 6 months]), and conversely longer ones being more prevalent (85% vs. 4% [for more than 24 months]). Whilst the Tajik sample had a higher prevalence of a violent index offence (63% vs. 38%), the proportion with previous violent crime convictions was less in the former (8% vs. 53%).

The reoffending rates also differed between the validation and derivation datasets. Incidence of the primary outcome (i.e. violent reoffending over the 1-year follow-up) was similar in both cohorts (15% [or 144 people released from prison in Tajikistan] vs 12% in Sweden). However, the rate of any reoffending was considerably lower amongst Tajik people released from prison (15% [147 individuals]) compared with their Swedish counterparts (44%). As a result of these differences, I made adaptations to the variable definitions prior to validating OxRec to account for the prevalence of predictors (see Appendix B Table B.2).

3.3.2 Model performance and updating

In the unadjusted model (i.e. simple validation), OxRec had good overall discrimination, with an AUC of 0.69 (95% CI 0.64–0.73). With regard to calibration, the performance of the validation model was poor (CITL = 1.76; Slope = 0.88), indicating an underestimation of risk across all risk deciles. That is, the expected number of outcome events (with respect to the model's prediction) was lower than the observed number of violent reoffending, suggesting that the validation model required recalibration.

Thus, I recalibrated the model as per study protocol, by which I updated the baseline risk and calculated a single multiplicative recalibration value. After adjusting the baseline risk and applying the recalibration value to the model, the validation model showed good calibration (CITL = 0.11; Slope = 1.00). The new estimates for the recalibrated model (i.e. baseline risk and multiplicative recalibration shape parameter) can be found in Table 3.2.

Moreover, I performed a selective re-estimation of coefficients for a single predictor (i.e. length of incarceration) to account for effect difference between the development and validation populations. This step was necessary as redefining the cut-offs for length of incarceration (due to considerable differences in prevalence) proved to be insufficient, which suggested that regression coefficients actually differed between the two settings (Appendix B Table B.5). This differential effect could be explained by the mass amnesty announced by the Tajik government in October 2019, where most people in prison were granted an official pardon, and thus released prior to the end of their sentence.

The revised model discrimination was indicated by an AUC of 0.70 (95% CI 0.65–0.75). Assuming a 15% risk cut-off to maximise the sum of sensitivity and specificity (and minimise the likelihood of false positives and false negatives) (Kaivanto, 2008), sensitivity was 60% (95% CI 0.52–0.68) and specificity was 65% (95% CI 0.62–0.69), whilst positive and negative predictive values were 23% (95% CI 0.19–0.28) and 90% (95% CI 0.88–0.93), respectively. This final model had a calibration slope of 1.09, and CITL was null. Calibration plots and ROC curves before and after model revision are shown in Appendix B Figure B.1, and Figures 3.1 and 3.2, respectively. Performance measures for the updated model are presented in Table 3.3.

3.4 Discussion

In this prospective cohort study of 970 people released from prison, I tested the performance of a risk assessment tool (OxRec), for the primary outcome of violent reoffending. 15% of participants reoffended within one year of release. OxRec showed good discriminative ability with an AUC of 0.70 (95% CI, 0.65–0.75). Updating the model resulted in good calibration (E:O = 1.00; CITL = 0.00; Slope = 1.09). To the best of my knowledge, this study is the first to validate a structured violence risk assessment tool in an LMIC.

Table 3.2: Recalibrated model formulae

	Model formulae	Baseline risk coefficient
Sweden	$1 - S^{exp(\sum \beta \times RF)}$	$S = 0.7992$
Tajikistan	$1 - S^{exp(0.8093 \times [-0.0348 \times 0.3075 + 0.0259 \times 0.39 + -0.0098 \times prison_{d2} + 0.5949 \times prison_{d3} + -0.1066 \times prison_{d4} + \sum \beta \times RF])}$	$S = 0.4708$

Note. β and RF refer to the model coefficients and risk factors presented in Fazel, Chang, et al. (2016), respectively. The multiples of 0.3075 and 0.39 are adjustments to allow for the immigrant status and neighbourhood deprivation variables being entirely missing in the validation study. '$prison_d$' corresponds to the length of incarceration (which was re-estimated in the Tajik cohort), and the multiples of -0.0098, 0.5949, -0.1066 are the updated coefficients for the variable.

Table 3.3: Summary of updated model performance

	Incidence of reoffending	AUC (95% CI)	Risk threshold	Sensitivity	Specificity	PPV	NPV
Violent reoffending (1 year)	15%	0.70 (0.66–0.75)	5%	99% (98–100)	8% (6–10)	16% (13–18)	99% (96–100)
			10%	88% (82–93)	37% (34–41)	20% (17–23)	94% (92–97)
			15%	60% (52–68)	65% (62–69)	23% (19–28)	90% (88–93)
			20%	41% (33–49)	81% (78–84)	27% (21–33)	89% (86–91)

Note. AUC = area under the curve; NPV = negative predictive value; PPV = positive predictive value.

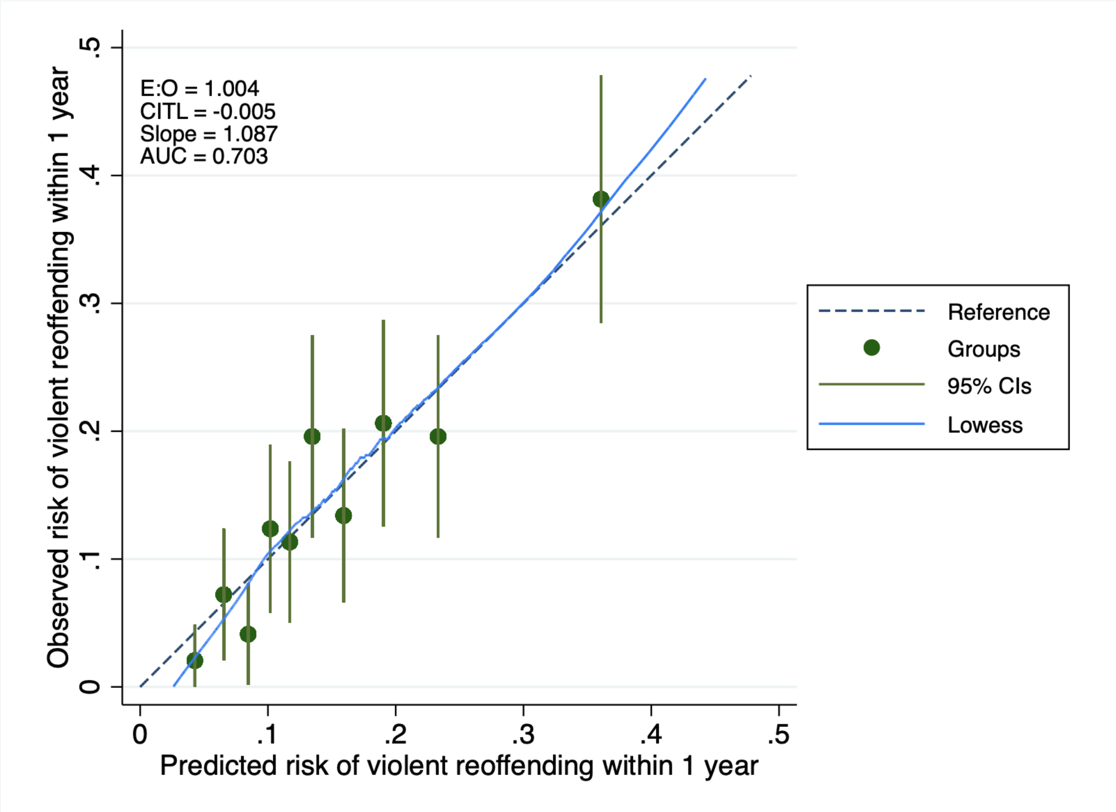


Figure 3.1: Calibration plot of the OxRec model performance in the Tajik cohort. AUC = area under the curve; CITL = calibration in the large; E:O = ratio of expected to observed outcomes.

3.4.1 Main findings

The observed performance of OxRec in this new setting indicated only a slight decrease in predictive performance, despite differences in social, cultural and economic contexts. In Sweden, the AUC was 0.76 for the same outcome. Country-specific variations, which tend to be particularly pronounced between HICs and LMICs, may explain differences in the distribution of participant characteristics, and in turn influence the performance of the model (Justice et al., 1999; Vergouwe et al., 2010).

The Tajik version of OxRec achieved similar levels of discriminative performance to other common risk assessment tools, all of which were developed in HICs. A recent systematic review of external validation studies of risk assessment tools for repeat offending (Fazel, Burghart, et al., 2022), which included 33 studies with nearly 600,000 participants, reported AUCs that ranged from 0.57 to 0.75. Another review, which focused on tools used in US-specific correctional settings to estimate recidivism

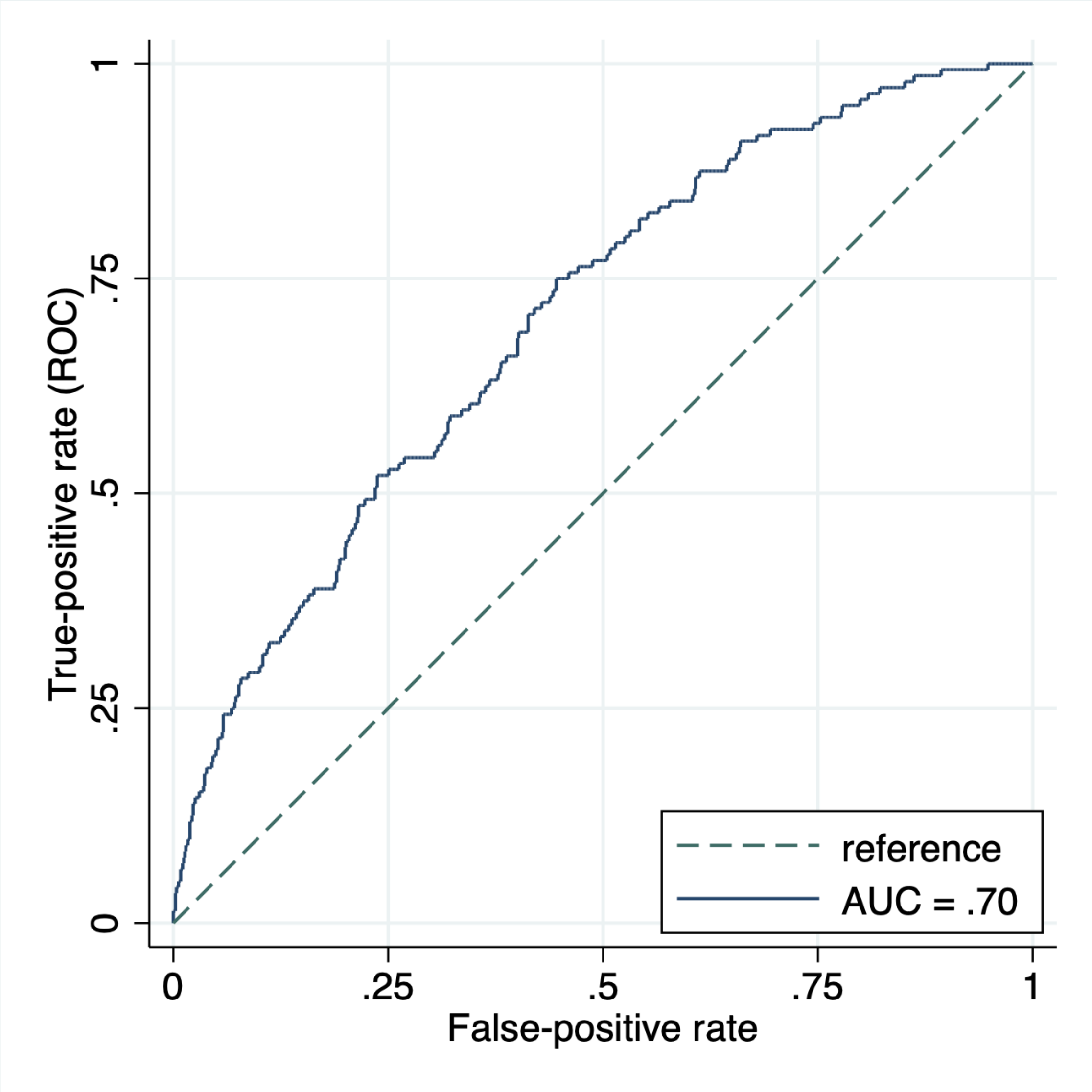


Figure 3.2: Receiver-operating characteristic curve for performance of the OxRec model in predicting violent reoffending outcome in the Tajik cohort within one year of release from prison. AUC = area under the curve; ROC = receiver operating characteristic curve.

risk, found similar estimates of predictive validity (Desmarais et al., 2016). However, included primary studies in these prior reviews failed to report key performance measures, such as other discrimination statistics (i.e. true and false positives and negatives) and calibration, which means more detailed comparisons with OxRec are not possible. Solely using AUCs to summarise and compare the prognostic accuracy of risk tools is uninformative, as it does not allow for the consideration of false negative and false positive predictions, which vary across risk thresholds, some of which may not be clinically relevant (Halligan et al., 2015; Mallett et al., 2012).

The discriminative ability of OxRec in the Tajikistan cohort was similar to that assessed in another OxRec validation study, which included both people in prison and individuals on probation from the Netherlands, as measured by the AUC (range 0.65–0.68 [for one and two-year violent reoffending outcomes]). The tool also required recalibration there, due to an overestimation of risk resulting from a lower incidence of violent reoffending (8% [Netherlands] vs 12% [Sweden]). The finding that the OxRec model replicated well in Tajikistan highlights the relevance of known risk factors—previously tested in a HIC context—for violence reoffending in an LMIC setting. The modifiable risk factors included in OxRec, such as mental illness and substance use disorders, are highly prevalent in people in LMIC prisons, and reflect considerable unmet mental-health needs (Mundt et al., 2018). Hence, they could be directly targeted in intervention programmes for people released from prison in resource-poor settings to reduce recidivism.

Furthermore, the missingness of two predictor variables (i.e. neighbourhood deprivation and immigrant status), did not materially alter the model's predictive performance. This finding is unsurprising considering that they are the weakest predictors (according to OxRec's risk factor weighting) (Fazel, Chang, et al., 2016). Immigrant status was found to be protective for recidivism in the Swedish studies, and therefore can be justified on ethical reasons as it might mitigate possible professional biases. However, the effect is likely to be different in LMICs.

3.4.2 Implications

I have validated a scalable, feasible tool to predict future repeat violent crime in people released from prison in a LMIC setting. With minor adjustments, including a straightforward recalibration, the OxRec tool can provide an effective way of identifying individuals at highest risk of violent reoffending, and in turn prioritise them for community interventions upon release from prison, such as substance use clinics. Such tools are needed, particularly in Tajikistan where the prison population is affected by the country's role in drug transit, which translates into high rates of injection drug use and related infections (i.e. HIV and tuberculosis) (Winetsky et al., 2014). This also has relevance for LMICs more broadly, where resources are limited due to low levels of investment in prison and community public health services (Ako et al., 2020). Even when available, considerable barriers to accessing resources exist for justice-involved individuals in LMICs, partly due to fear of stigmatisation, lack of treatment in community settings and limited linkage between services (Komalasari et al., 2021).

Moreover, the performance of OxRec in terms of obstacles to external validation, such as outcome incidence, case-mix (i.e. population characteristics) and predictor effects, highlights its potential generalisability to other LMIC settings. Future research could investigate novel risk factors that might be unique to LMICs, although this may involve the development of new models.

As there is a paucity of research on violence risk assessment in LMICs, this study provides one approach. I would like to highlight several challenges to risk prediction in LMICs, encountered in the process of testing the predictive ability of OxRec in Tajikistan. These should be considered prior to undertaking subsequent validations. Most assessment methods employed in HICs necessitate the involvement of trained research professionals. However, such expertise is scarce and typically too costly for LMIC settings (Rodriguez & Roehr, 2020). Prison health-care services in LMICs are more than often under-resourced. This means that there is a lack of resources for conducting research, and thus data collection, especially when this involves a large number of predictors, given the absence of electronic records, or even readily available registries (Tirupakuzhi Vijayaraghavan et al., 2020). Some data are partially missing, or reported inconsistently, whilst others are simply not available in LMICs (Haniffa et al., 2018). Therefore, substantial model adaptations and refinement are required during the validation process.

3.4.3 Strengths and limitations

This validation study has several strengths. I present key performance measures of discrimination (i.e. true and false positives and negatives) for various risk thresholds, in addition to the AUC. Those are often overlooked in validation studies, despite being important in terms of potential consequences for justice-involved individuals and the wider society, as they are likely to inform decisions relating to rehabilitation and public safety (Fazel et al., 2019). The study design is another strength. Prospective cohort studies, as opposed to retrospective ones, offer the advantage of compiling more reliable information regarding exposures, confounders, and end points, and allow for the estimation of outcome incidence (X. Wang & Kattan, 2020).

Some limitations should be noted. Several adaptations were made to variable definitions, and two predictors, neighbourhood deprivation and immigrant status, were omitted. I relied on self-reported information for variables relating to mental illness and substance use disorders, which may introduce report bias especially in LMICs due to stigma (Komalasari et al., 2021; Mascayano et al., 2015), whereas diagnosed conditions were used in the development study. However, it was not possible to rely on diagnoses in Tajikistan, and likely in other LMICs, due to the relative lack of public health care services. Therefore, I determined cut-off scores for self-reported predictors bearing in mind reflect prevalences in the development sample. I could not stratify results by sex (as assigned at birth) as the number of women included in the Tajik sample was insufficient for a validation study.

Moreover, it was not possible to integrate the most recent recommendations on the estimation of required sample size for the external validation of risk prediction models for survival outcomes, as these were published after analyses were conducted. Future validation studies should be devised using newly proposed sample size estimators, given that these not only take into account outcome incidence, but also model strength as measured by key calibration measures (i.e. AUC, calibration slope and CITL) (Pavlou et al., 2021).

3.5 Conclusion

3.5.1 What this chapter adds

The contribution of this chapter to literature is two-fold. Firstly, I have established that OxRec—a scalable risk assessment tool—could be used in Tajikistan and potentially other LMICs, to improve the consistency, transparency and accuracy of risk assessment, and linkage to community services. Secondly, findings suggest that existing risk prediction models for predicting repeat offending and violent crime in justice-involved individuals, which have been developed and validated in HICs, could be applicable in LMICs pending rigorous validation studies to ascertain their transportability to these new settings. However, its relevance goes beyond psychiatry and prison health research, as it provides another example of prediction models developed in HICs that can be adapted to LMICs. This adds to successful applications in other medical fields, such as neurology (e.g. Stephan et al., 2020), gastroenterology (e.g. Levine et al., 2016), obstetrics (e.g. Ukah et al., 2017).

Contribution statement

This chapter is based on the published work: *Beaudry, G., Yu, R., Alaei, A., Alaei, K., & Fazel, S. (2022). Predicting violent reoffending is individuals released from prison in a lower-middle-income country: a validation of OxRec in Tajikistan. Frontiers in Psychiatry, 13(805141), 1–9.* I conducted the analyses, and prepared the tables and figures under the supervision of Dr Rongqin Yu. I also wrote the first draft of the manuscript and implemented the contribution of the co-authors and external reviewers up to final publication. I am grateful to Dr Thomas R. Fanshawe for providing feedback on the statistical analyses.

3.5.2 Next chapter

In the next chapter of this thesis (Chapter C), I will employ the same methodological procedure as described in the current one to externally validate OxRec in England. Findings will enable comparison of model performance between the two cohorts (Tajikistan vs. England).

4 OxRec external validation in England

Contents

4.1 Introduction

In the previous chapter (Chapter 3), I described the first validation study of a structured violence risk assessment tool in a LMIC. Specifically, I externally validated a scalable model for predicting violent reoffending in people released from prison (OxRec)—which was developed in Sweden—using data from a cohort of people who had previously experienced incarceration in Tajikistan. The OxRec model demonstrated good performance in multiple measures, suggesting that it could be used in Tajikistan to help prioritise interventions for people who are at high risk of violent reoffending after incarceration and screen out others who are at lower risk of violent reoffending. Findings also indicated that the use of validated risk assessment tools in LMICs could improve risk stratification and inform the development of future interventions tailored at modifiable risk factors for recidivism, such as substance use and mental health problems. The objective of the present chapter is to validate OxRec in another HIC (England) than the one in which it was derived (Sweden) and if necessary, to adjust or update the model for this new setting. As a secondary objective, the predictive performance of OxRec in Tajikistan and England will be compared, as will be the generalisability of the model in these two countries.

As described in Chapter 1, predicting future outcomes, such as reoffending, has increasingly become a priority in the criminal justice system (Harcourt, 2006). Therefore, predictive modelling approaches are employed in many high-income countries as part of efforts to save costs and improve efficiency in criminal prosecution (Van Ginneken, 2019). Their applications are widespread and vary from informing decisions on sentencing, parole, release, probation and treatment, particularly for mental health problems and substance misuse (Fazel, 2019). Recent meta-analytic evidence suggests that more than 200 such risk assessment tools are currently in use (Singh et al., 2014), however most of these are yet to be validated in new settings or individuals. In a recent review of external validation studies of 11 common risk assessment tools used at sentencing, only 36 studies were identified (Fazel, Burghart, et al., 2022), thus highlighting the imbalance between model development and external validation efforts to date (Ramspek et al., 2021).

Very few prediction models have been validated in jurisdictions different to the one in which they were developed. Predictive performance in validation samples has often been suboptimal in terms of AUC. Reporting of model performance (using discrimination and calibration measures also tends to be poor (Fazel, Burghart, et al., 2022). The majority of studies only report the AUC, albeit having limited interpretation potential as a standalone metric considering that it measures performance across a whole range of cut-offs, most of which are not clinically relevant (Mallett et al., 2012). More detailed reporting of models (beyond the AUC)—including information on false positives and negatives, and calibration measures—is essential for evidence-based clinical decision making (Gulati et al., 2022), yet it is far from being the norm in prediction modelling research. For instance, one such tool—the Offender Assessment System (OASys), which has been revised since its inception into the OASys Violence Predictor (OVP)—is routinely used by probation officers in England and Wales to assess individual needs and future risks (P. Howard, 2006), despite limited information on its predictive ability. Thus far, the only disclosed performance metrics pertain to discrimination (Debidin, 2009; P. Howard & Dixon, 2011; P. D. Howard & Dixon, 2012). Detailed reporting of general performance and calibration measures is required to enable adequate comparison to other violence risk assessment tools (Collins et al., 2015; Steyerberg, 2009). Furthermore, the vast majority of tools are never implemented in practice, and despite this, research efforts are mostly direct at developing new prediction models rather than evaluating and, validating or updating existing ones (Collins et al., 2014).

The OxRec tool was developed in recent years to address some of these limitations (see Chapters 1 & 3 for more detail). In this chapter, I conduct the validation of OxRec to estimate one- and two-year violent reoffending risk in people released from prison in England. If OxRec demonstrates satisfactory predictive performance in this new setting, it could be used to improve identification of individuals at high risk of violent reoffending, and inform decision-making processes in the criminal justice system. This would also provide more evidence for the transportability of OxRec to new settings, adding to promising validation results in the Netherlands and Tajikistan. There is evidence suggesting that the test of the model's generalisability is stronger the more different these other settings are to that of the derivation study (Moons et al., 2012).

4.2 Methods

4.2.1 Study design and participants

This validation study used data on criminal records that pooled information from people released from prison in England aged 18 years and older or treated as adults by the law. The study sample included persons released from prison between April 1, 2017 and March 31, 2018 in one English region of around one million population that includes both metropolitan and non-metropolitan areas. People released from prison were censored at the earliest date of the reoffence or the follow-up end date (within 24 months). The sample size was based on prognostic modelling guidelines according to which at least 100 events (and 100 nonevents) are required for validation studies (Vergouwe et al., 2005). This benchmark ensure that there is sufficient statistical power for predictive performance to be assessed in a reliable manner. The model becomes more sensitive to outliers and high variance as the sample gets smaller, which in turn increases the risk of overfitting (Steyerberg, 2018).

Individuals entering the Criminal Justice Information Service are informed about secondary usage of data, thus the standard requirement of written informed consent was waived. I used existing routinely collected police data. Ethical approval was granted by the Central University Research Ethics Committee (CUREC) of the University of Oxford (R44562/RE001). This study complies with the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis statement (Collins et al., 2015).

4.2.2 Outcomes

I obtained information on reoffending outcomes from the NICHE Record Management System (the system that the police force uses to record offences reported to and investigated by the force). The following outcome-related data were recorded: the person was suspected of committing a violent offence by the police force, had been convicted for a violent offence, had not reoffended or had been lost to follow-up. At one- and two-years following release from prison, the outcomes were described in binary terms (i.e. 'violent reoffence' vs. 'no violent reoffence'). The definition of violence was a standard one used by the police force that included two categories of 'assault violence or threat of violence' (i.e. any interpersonal violence and violent threats) and 'rape and penetrative sexual offences' (i.e. any contact sexual crimes).

4.2.3 Predictor variables

I examined the predictor variables based on those already included in the original version of OxRec. Several modifications to variable definitions were required to adapt OxRec to the English regional context (Appendix C Table C.1). Notably, sex (assigned at birth) was replaced by gender. Immigrant status was based on nationality, instead of first- or second-generation immigrant status in the development sample. Neighbourhood deprivation was calculated using the Index of Multiple Deprivation (IMD; 2019), which is the official measure of deprivation for small areas in England. More specifically, the last valid residential or family home address before the release date was linked with the Lower-layer Super Output Areas (LSOA) to determine the relevant IMD decile (Noble et al., 2019). Data on clinical variables (i.e. alcohol and drug use disorder, any mental disorder) were inferred from warnings recorded by police officers and custody staff in the NICHE Record Management System—rather than formal diagnoses according to the International Classification of Diseases 10th Revision (ICD-10). I could not include the following predictors, as they were not available in the English data set: highest education, disposable income and any severe mental disorder.

4.2.4 Statistical analysis

Baseline characteristics were detailed by counts and percentages for categorical variables, and as median (IQR) for continuous variables, and compared to those of the Swedish derivation sample. The prediction rule was validated using an incremental strategy (Steyerberg, 2009; Su et al., 2018), first adapted to validate OxRec in the Netherlands (Fazel et al., 2019), and subsequently used for another validation study in Tajikistan (see Chapter 3). This approach involves three steps (1) simple validation, according to which the original prediction model (all model coefficients at their original value, including the baseline risk estimate) are applied to the new data set; (2) updating the baseline risk and calculating a multiplicative recalibration value; (3) performing a selective reestimation of coefficients for individual predictors. Each phase evaluates prognostic performance, and progression to the next step is only necessary in the event of poor performance (Steyerberg, 2009). The full validation protocol (including recalibration with a Cox proportional hazards model) has been detailed in the previous chapter (Chapter 3) and has also been published online (Fazel et al., 2019).

I briefly discussed the implications of missing values for validating a prediction model in Chapter 3. I used a straightforward mean imputation method to mitigate the impact of some predictor variables being entirely missing from the Tajik validation data set. By assigning an average value (based on that of the Swedish development cohort) to these missing variables, their effect was essentially incorporated into the model intercept (Fazel et al., 2019). The current validation study required a more complex approach to dealing with missing data, as missing values arised in several predictors for some individuals.

I started by investigating potential missingness patterns in the English validation cohort to select the most appropriate imputation method based on the underlying mechanism assumption. Possible missing data mechanisms for predictors include (1) missing completely at random (MCAR); (2) missing at random (MAR); and, (3) missing not at random (MNAR) (Little & Rubin, 2020; Rubin, 1976; van Buuren, 2012). MAR provides that missing values are only associated with values of other predictors in the data set, rather than those of predictor itself or unmeasured variables. With an MCAR pattern, missing values are typically due to random administrative error, and this may be tested formally. Thus, this pattern could be ruled out on the basis of using already collected data from comprehensive police databases based on individual criminal records. The final potential mechanism of missing data (MNAR) is problematic in that missingness is contingent on the true values of the predictor or other unobserved variables (Steyerberg, 2009).

The observed data alone is insufficient to differentiate between MAR and MNAR; however, the MAR assumption can be supported by collecting additional independent variables in the data set (that are not predictors in the model), and considering them throughout the imputation process (White et al., 2011).

Imputation methods can be used to estimate plausible values for substitution of missing data. Once the data set is "completed," common statistical analyses can then be undertaken. There is a wide range of imputation methods, including simple (e.g. mean substitution, whereby missing values are replaced with the mean of observed data) and more advanced techniques (i.e. multiple imputation). Simple imputation methods have the disadvantage of inappropriate handling of between-subject variance (White et al., 2011) and not considering potential predictor value correlations. Conversely, multiple imputation takes into account all available data (both from complete and incomplete cases) to construct several imputed data sets in which the missing values are replaced (Steyerberg, 2009). This particular approach holds the MAR assumption, and it is consistent with current best practice for handling missing data in prognostic modelling research (Marshall et al., 2009).

For this validation study, data were assumed to be missing at random (MAR). Thus, multiple imputation by chained equations (MICE) could be used to replace missing values for the following predictors: immigration status, length of incarceration, violent index offence, civil status and employment. Twenty imputations were carried out according to recommended practice (Steyerberg, 2009; White et al., 2011). I also averaged out the predictors that were entirely missing (i.e. highest education, disposable income and any severe mental disorder), by assigning all subjects their average value from the derivation sample. More detail on the rationale for this simple imputation method can be found in Chapter 3.

I examined several model performance indicators across the imputed data sets (using Rubin's rules; Snell et al., 2018; Wood et al., 2015) to determine the predictive ability of the model in terms of accuracy (overall performance), discrimination (the model's ability to discriminate between individuals who have and do not have the outcome), and calibration (the level of agreement between observed and expected outcomes) (Steyerberg, 2009). These indicators included the Brier score (for accuracy); the AUC, as well as sensitivity, specificity, PPV and NPV (for discrimination); and, the calibration slope and CITL (for calibration) (Fazel, 2019; Steyerberg, 2009). By comparing the distributions of actual and imputed values for all metrics, I was able to verify the adequacy of imputations (Ganna & Ingelsson, 2015). More detail on the aforementioned measures of predictive performance are presented in Chapter 3. All analyses were conducted using Stata software version 17 (StataCorp, 2017).

4.3 Results

4.3.1 Baseline characteristics

The validation cohort included 1,770 people released from prison in England. The median (IQR) age was 33 (27–40) years and male individuals accounted for 92% of the sample. The proportion of missing data on predictors varied from $< 5\%$ (immigrant status, length of incarceration and violent index offence); 9% (civil status); to 15% (employment status). Most baseline characteristics were similar between the validation (England) and the original derivation (Sweden) cohorts (Table 4.1). However, the proportion of individuals with a history of violent crime (prior to the index offence) was slightly higher in the validation sample than in the derivation one (62% vs. 53%). Compared with previously incarcerated persons in Sweden, those in England more often were incarcerated for longer periods of time for their index offence (31% vs. 4% for ≥ 24 months). A higher proportion of employment (63% vs. 25%) and mental health problems (35% vs. 22%) (including drug misuse [36% vs. 23%]) was observed in the sample from England, whereas alcohol misuse was more common amongst the original Swedish sample (22% vs. 14%).

Follow-up data were obtained from all individuals who were included in the study (Appendix C Table C.2). Base rates of violent reoffending (i.e. including both suspected and convicted for violent crime for this validation) for the two time points (1 and 2 years) were higher in the validation cohort than in the derivation cohort. The primary outcome, violent reoffending, occurred in 31% (550 of 1,770; one-year

Table 4.1: Baseline characteristics of the English sample compared with those of the Swedish sample

Variable	English sample (n = 1,770)		Swedish sample (n = 37,100)
	Summary	Missing data	
Sex			
Male	1,622 (92%)		93%
Female	148 (8%)		7%
Age	Median 33		Median 36
	IQR 27–40		IQR 27–46
Immigrant status	115 (7%)	1 (< 1%)	31%
Length of incarceration			
< 6 months	658 (37%)		69%
6–12 months	253 (14%)		16%
12–24 months	269 (15%)		10%
≥ 24 months	545 (34%)		4%
Violent index offence	597 (34%)	18 (1%)	38%
Previous violent crime	1,095 (62%)		53%
Civil status			
Other	321 (18%)	150 (9%)	35%
Unmarried	1,299 (73%)		65%
Education			
< 9 years	Not available		48%
9–11 years			46%
≥ 12 years			6%
Employment	1,116 (63%)	258 (15%)	25%
Income			
Negative	Not available		1%
Zero			6%
Low			53%
Medium			40%
High			1%
Neighbourhood deprivation	0.67 (0.13–1.04)		0.39 (-1.18–1.47)
Alcohol use disorder	249 (14%)		22%
Drug use disorder	631 (36%)		23%
Any mental disorder	627 (35%)		22%
Any severe mental disorder	Not available		3%

Note. Data are median (IQR) or n (%).

follow-up) and 43% (765 of 1,770; two-year follow-up) of the sample from England compared with 12% (one year) and 21% (two years) in the sample from Sweden.

4.3.2 Model performance and recalibration

When refitting the OxRec model in the validation data, the discriminative ability remained at a similar level as at derivation (AUCs of 0.75 [one year] and 0.76 [two years] in the Swedish sample). AUCs for violent reoffending at one and two years were 0.71 (95% CI 0.69–0.74) and 0.71 (95% CI 0.68–0.73), respectively. I used cut-off scores that were close to the baseline rates (31% at one-year; 43% at two-years) to calculate values for sensitivity, specificity, NPV, and PPV. For risk of violent reoffending at one year (assuming a 30% risk cut-off), sensitivity was 74% (95% CI 0.70–0.78) and specificity was 59% (95% CI 0.56–0.62), whilst positive and negative predictive values were 45% (95% CI 0.42–0.48) and 83% (95% CI 0.81–0.86), respectively. At 2 years, sensitivity was 77% (95% CI 0.74–0.80) and specificity was 54% (95% CI 0.51–0.58), using a 40% risk cut-off. Positive and negative predictive values were 56% (95% CI 0.53–0.59) and 76% (95% CI 0.73–0.79), respectively. These discriminative performance measures are presented for additional risk cut-offs in Table 4.2.

As for calibration, performance parameters suggested miscalibration for both time points (CITL = 1.72; Slope = 0.98 [one year]; CITL = 1.73; Slope = 0.92 [two years]), with observed violent reoffending probabilities in England being systematically higher than expected. I updated the baseline survival function and recalibrated the linear predictor to align the predicted and observed survival probabilities for all risk deciles (Table 4.3). This step improved OxRec's accuracy (in terms of calibration) with respect to the new English regional setting. The revised model showed good calibration with a calibration slope of 1.24 (one year) and 1.38 (two years), and CITL being null for both time points. Despite this, calibration plots indicated a slight overestimation of violent reoffending risk amongst lower risk deciles and a slight underestimation of risk in the higher deciles. The effects of predictors were similar in the development and validation samples, thus no reestimation of the

Table 4.2: Summary of updated model performance

	Prevalence	AUC (95% CI)	Risk threshold	Sensitivity	Specificity	PPV	NPV
Violent reoffending (1 year)	31%	0.71 (0.69–0.74)	20%	92% (89–94)	31% (29–34)	38% (35–40)	89% (86–92)
			30%	74% (70–78)	59% (56–62)	45% (42–48)	83% (81–86)
			40%	44% 40–49)	82% (80–84)	52% (48–57)	77% (74–79)
Violent reoffending (2 years)	43%	0.71 (0.68–0.73)	30%	93% (91–95)	27% (24–30)	49% (47–52)	83% (79–87)
			40%	77% (74–80)	54% (51–58)	56% (53–59)	76% (73–79)
			50%	49% 46–53)	77% (74–79)	62% (58–65)	67% (64–69)

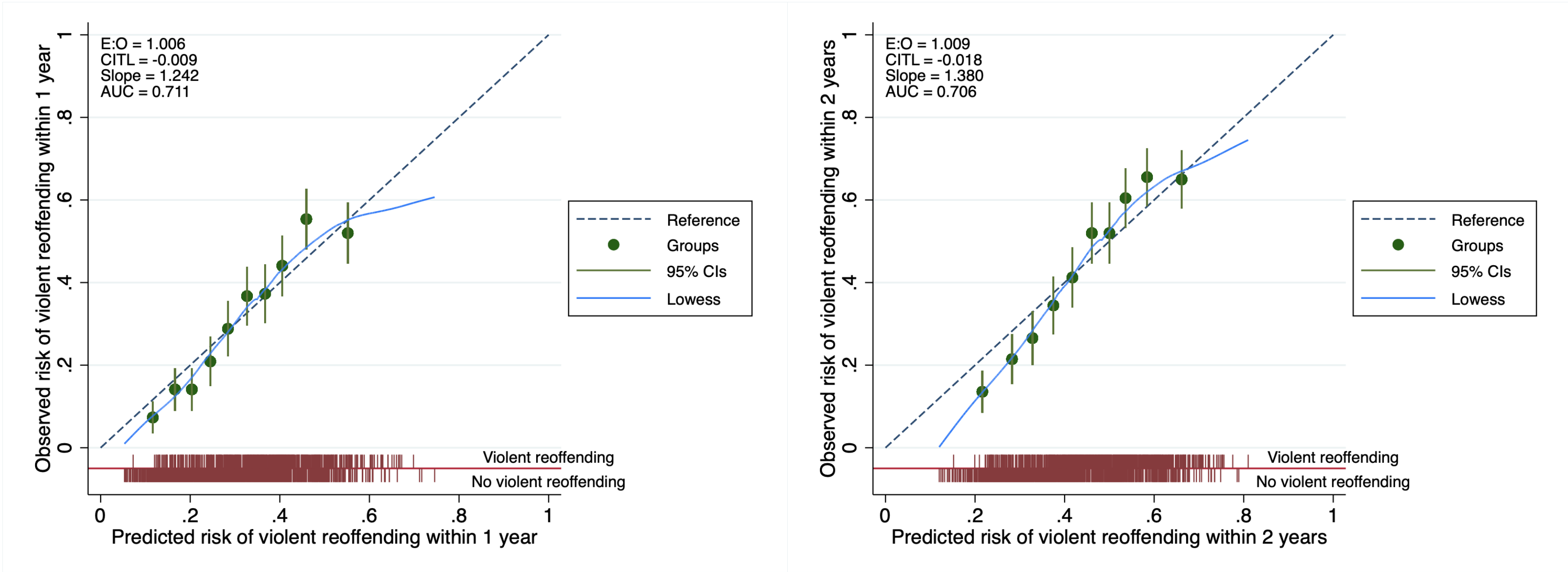
Note. AUC = area under the curve; NPV = negative predictive value; PPV = positive predictive value.

original coefficients was required (Appendix C.4). The ROC curves and calibration plots for the updated OxRec model are shown in Figures 4.1 and 4.2.

Table 4.3: Recalibrated model formulae

Sweden	Model formulae	Baseline risk coefficient
Violent reoffending (1 and 2 years)	$1 - S_t^{exp(\sum \beta \times RF)}$	$S_1 = 0.7992$ $S_2 = 0.6775$
England		
Violent reoffending (1 year)	$1 - S_1^{exp(0.6745 \times [-0.1838 \times 0.4263 - 0.4282 \times 0.0569 + 0.5251 \times 0.0523 + 0.5176 \times 0.4903 + 0.3712 \times 0.3738 + 0.4509 \times 0.0109 + 0.0953 \times 0.0089 + \sum \beta \times RF])}$	$S_1 = 0.4643$
Violent reoffending (2 years)	$1 - S_2^{exp(0.6745 \times [-0.1838 \times 0.4263 - 0.4282 \times 0.0569 + 0.5251 \times 0.0523 + 0.5176 \times 0.4903 + 0.3712 \times 0.3738 + 0.4509 \times 0.0109 + 0.0953 \times 0.0089 + \sum \beta \times RF])}$	$S_2 = 0.3509$

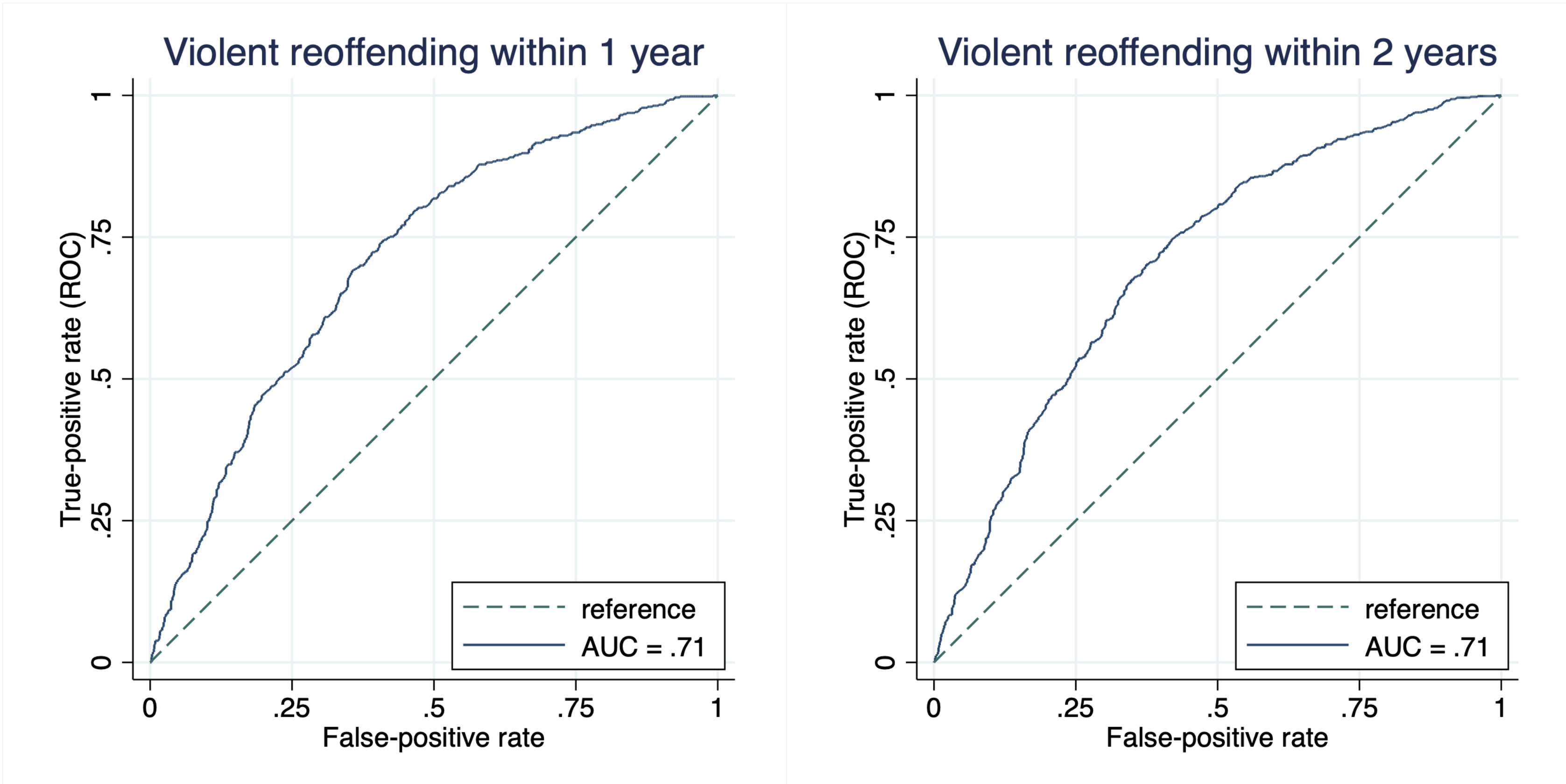
Note. β and RF refer to the model coefficients and risk factors presented in Fazel, Chang, et al. (2016). The following multiples are adjustments to allow for some predictors being entirely missing in the validation study: highest education (0.4263 and 0.0569); disposable income (0.0523, 0.4903, 0.3738 and 0.0109); any severe mental disorder (0.0089).



(a) 1-year violent reoffending

(b) 2-year violent reoffending

Figure 4.1: Calibration plots of the OxRec model performance in the English cohort. AUC = area under the curve; CITL = calibration in the large; E:O = ratio of expected to observed outcomes.



(a) 1-year violent reoffending

(b) 2-year violent reoffending

Figure 4.2: Receiver-operating characteristic curve for performance of the OxRec model in predicting violent reoffending outcome in the English cohort. AUC = area under the curve; ROC = receiver operating characteristic curve.

4.4 Discussion

In this study, I externally validated OxRec in a cohort of 1,770 people released from prison in England between April 2017 and March 2018, of which the one- and two-year violent reoffending rate were 31% and 43%, respectively. The model yielded good discrimination and calibration. This work builds on previous validation studies of OxRec in the Netherlands and Tajikistan, and further demonstrates its generalisability and the accuracy of its predictions in new settings. The performance of OxRec was confirmed in this new sample, suggesting that it could be used to improve current risk assessment practices for violent reoffending in the criminal justice system in England.

OxRec accurately identified individuals at high risk of violent reoffending using cut-offs of 30% for 1-year risk and 40% for 2-year risk (i.e. 74% and 77% of whom did reoffend within 1 and 2 years, respectively). These individuals could be prioritised for additional support on release, such as psychological interventions directed at modifiable risk factors (e.g. substance misuse), given the limited resources and the likelihood that they will benefit. The tool also accurately identified individuals at low risk of violent reoffending based on the aforementioned cut-offs (i.e. 59% and 54% for one- and two-year follow-up periods), although specificity was lower than sensitivity. This could inform decarceration efforts, whereby individuals in this low risk group could be released with adequate supervision and treatment in the community (Fazel, Burghart, et al., 2022). Used in conjunction with professional judgement, high-quality risk assessment tools such as OxRec have the potential to reduce economic and social costs in the criminal justice system, allowing for jurisdictions to focus on establishing policies that assist people with convictions in reintegrating into society.

The OxRec tool had similar performance for identifying violent reoffending outcomes in the English validation cohort as compared to that in the Swedish derivation cohort. I also found similar associations of risk factors included in OxRec with violent reoffending in the derivation and validation cohorts (Fazel, Chang, et al., 2016). The base rate of violent reoffending in the English sample was nearly double that of the Swedish sample (i.e. 31% vs. 12% [for 1 year] and 43% vs. 21%

[for 2 years]), and thus predicted probability calculated with OxRec underestimated the observed probability of violent reoffending, prior to model updating. This difference is likely due to adaptations made to the outcome operationalisation (i.e. conviction rate in the Swedish sample vs. suspected and conviction rate in the English sample), as recidivism statistics are similar in both countries according to a recent systematic review (Yukhnenko et al., 2019).

The findings of this study further suggest that OxRec is transportable across different populations and geographical settings, as evidenced by good discrimination and calibration in external validation data sets from England, Tajikistan (Chapter 3) and the Netherlands (Fazel et al., 2019). As indicated by the AUC, the model achieved similar levels of discrimination in the English sample (0.71) to that of the Tajik sample (0.70). However, as explained in Chapter 1 and earlier in the current chapter, the AUC only provides a limited view of the discriminatory power. AUCs are independent of the outcome base rate, thereby providing no indication as to how the model influences results for specific individuals, and in turn, not enabling detailed comparison between samples (Mallett et al., 2012; Van Calster et al., 2019). When considering more comprehensive measures of discrimination (for one-year violent reoffending, and by using the outcome incidence as the threshold), sensitivity and PPV were higher in England (Table 4.2), whereas specificity and NPV were higher in Tajikistan (Table 3.3). This finding has important implications for the intended use of the model (at this stage); OxRec might be more useful for correctly identifying individuals at high risk of violent reoffending in England, and conversely for ruling out those at low risk of violent reoffending in Tajikistan. As for calibration, CITL was similar across both samples, but there was a higher calibration slope in the English cohort (1.24) than in the Tajik cohort (1.09), suggesting more case-mix differences in England. However, the model updating process was more straightforward in England than in Tajikistan, where only the calculation of a single recalibration parameter was required (second step of the aforementioned validation protocol) (Fazel et al., 2019). By contrast, OxRec Tajikistan required the re-estimation of some model coefficients (i.e. length of incarceration), equivalent to the third step of the pre-specified validation protocol (Chapter 3).

In comparison with OVP—currently being used by the National Offender Management Service in England and Wales for risk and needs assessment—the OxRec tool achieved similar levels of predictive performance when considering solely the AUC (0.71 [OxRec] vs. 0.75 [OVP]) (P. D. Howard & Dixon, 2012). However, other key components of OVP's predictive performance are poorly reported or simply lacking. Measures of discrimination beyond the AUC (i.e. sensitivity, specificity, PPV, NPV) were calculated based on arbitrary classification risk thresholds, solely selected to match the distribution of another tool's categories (i.e. the V scale of Risk Matrix 2000 [RM2000/V]) for comparison purposes (Thornton et al., 2003). Calibration performance was not reported, despite being essential to ensure that predictions are not misleading (Van Calster et al., 2019), and recommended by the TRIPOD guidelines for prediction modelling studies (Collins et al., 2015). By contrast, OxRec's external performance was evaluated using pre-specified risk thresholds (for discrimination) and appropriate measures and visualisations (for calibration). OxRec also has the advantage of including fewer predictors than OVP, most of which are common and routinely collected, thereby allowing for its direct application and use in everyday practice. Hence, it can be easily calculated (in less than ten minutes) and incorporated into existing risk assessment practices by probation officers. Further, OVP has also yet to be validated independently (P. D. Howard & Dixon, 2012), and there is evidence that predictive performance tends to be overestimated when tool developers are involved in subsequent validation studies due to authorship bias (Fazel, Burghart, et al., 2022; Singh et al., 2013). Another key strength of this validation study is that it used data from the NICHE Record Management System, as do most of the UK police forces. This suggests that the validated tool could prove to be highly transferable to other criminal justice services in the UK.

Recent advances in criminal risk assessment and precision psychiatry, particularly in big data techniques (e.g. the Wisconsin's Correctional Offender Management

Profiling for Alternative Sanctions [COMPAS] tool), have favoured increased predictive power over result interpretability (Fusar-Poli et al., 2019; Meehan et al., 2022). The main strength of this study is that it combines accuracy of probabilistic predictions with a relatively simple prediction tool. The contribution of predictor variables to the outcome is clear, and their relationship is easily interpretable. Another issue that has emerged from prediction modelling research is the lack of transparency (Chekroud et al., 2021). Detailed reporting of performance metrics, allowing for the additional scrutiny that any tool intended for use in the criminal justice system should face, is often lacking (Meehan et al., 2022). In this study, I provided multiple performance indicators for independent examination, critical appraisal and reproducibility of the model and the methods used to validate it. Transparency in model development and validation is critical given the possible ramifications for justice-involved individuals and public health and safety, and to ensure a fair criminal justice system (Pencina et al., 2020).

Several limitations should be noted. First, since the NICHE database includes information on both crimes solved and those under investigation, this meant that I had to rely on a proxy for the outcome (i.e. violent reoffending). Another validation study with more clearly defined outcome data (i.e. only convictions for violent crime) is therefore recommended (Pencina et al., 2020). Operationalising violent crime as a conviction, as opposed to an arrest, is expected to lessen the likelihood of systemic biases in the estimation of violence among formerly incarcerated individuals. Second, some predictors were entirely missing from the English data set (i.e. highest education, disposable income and any severe mental disorder), whilst others (i.e. clinical risk factors) were insufficiently captured. To mitigate the impact of missingness, I employed MICE (for partially missing predictors), by which plausible values were derived based on other (observed) predictor values (Harel et al., 2018; Janssen et al., 2010). As for entirely missing predictors, although they had small effects on the predicted outcomes in the original model, all participants were assigned an average value (equivalent to the prevalence in the derivation study), which is the same as integrating its effect into the estimate of baseline risk (Fazel et al., 2019; Held et al., 2016). To reduce the possible effects of missing data, future research should strive for linkage of nationwide population-based registers, such as those found in the Nordic countries (Maret-Ouda et al., 2017), although access to such data is often costly and highly restricted. Given that the optimal approach to missing values is to guarantee that no data are absent (Steyerberg, 2009), this could be done using registry data as these tend to be comprehensive and reliable. Finally, this study provides no practical guidance on how OxRec could be utilised as a decision-support tool (both in custodial and community settings) to help reduce violent reoffending. A feasibility and clinical impact study, in which the feasibility of implementing OxRec in a given custodial setting in England, and its impact on current practices, reoffending outcomes and cost effectiveness would be evaluated (ideally using a cluster randomised trial design) (Labarere et al., 2014). This is required to determine how the risk prediction tool could be integrated effectively into existing operational systems and legal workflows, and whether modifiable risk factors in OxRec could be effectively targeted for treatment (Mudumbai & Rashidi, 2021).

4.5 Conclusion

4.5.1 What this chapter adds

The aim of this chapter was to directly address the second research question of this thesis by externally validating a prognostic model for violent reoffending in people released from prison (OxRec) based on an English regional cohort. The main contribution of this study is establishing that OxRec has satisfactory predictive performance and allows for stratification of individual risk of violent reoffending within 1 and 2 years after release from prison in England. This has similar implications for practice as the previous validation study in Tajikistan (Chapter 3), whereby OxRec could be used to supplement professional judgement and facilitate linkage to care following release from prison. Individuals in the high risk group could potentially benefit the most from empirically supported interventions, given limited resources in the criminal justice and health services. The use of high-quality risk assessment tools has the potential to provide more accurate and consistent evidence-based decision-making, improve policy-making in the criminal justice system, and also save costs due to time efficiency and reduced violent reoffending.

Contribution

I conducted the analyses, and prepared the tables and figures under the supervision of Dr Rongqin Yu. I also wrote the first draft of the manuscript and will be implementing the contribution of the co-authors and external reviewers up to final publication.

4.5.2 Next chapter

The next chapter (Chapter 5) will focus on the third and final research question of this thesis, which examines the effectiveness of psychological interventions in prison to reduce recidivism. Specifically, I will provide a comprehensive synthesis of the most evidence-based treatments, by limiting the scope of the systematic review and meta-analysis to RCTs.

This chapter is adapted from:

Beaudry, G., Yu, R., Perry, A. E., & Fazel, S. (2021). Effectiveness of psychological interventions in prison to reduce recidivism: a systematic review and meta-analysis of randomised controlled trials. *Lancet Psychiatry*, *8*(9), 759–773.

5

Psychological interventions in prison to reduce reoffending

Contents

5.1 Introduction

The systematic review and meta-analysis described in Chapter 2 established that psychiatric disorders are disproportionately prevalent in detained adolescents. As explained in Chapter 1, severe mental illness is associated with increased risk of general and violent reoffending amongst people released from prison (Fazel, Chang, et al., 2016), and pharmacological treatment has shown some beneficial effects for reducing risk (Chang et al., 2016). Scalable, transparent and evidence-based risk-assessment tools such as OxRec could assist in identifying individuals who are most likely to benefit from effective interventions (see Chapters 3 & 4). However, the current body of knowledge on the effectiveness of psychological treatment for other modifiable risk factors in reducing this adverse outcome is yet to be systematically synthesised. Thus, the objective of this chapter was to determine whether widely implemented psychological interventions for people in prison are effective to reduce recidivism after release.

Various psychological interventions have been implemented in prisons and jails to improve outcomes for people who have experienced incarceration, and in particular reduce reoffending. Some reviews have suggested that cognitive-behavioural therapy (CBT) programs are amongst the most effective, with meta-analyses reporting risk reductions of 20–30% in recidivism (Henwood et al., 2015; Landenberger & Lipsey, 2005; Lipsey et al., 2007; Lösel & Schmucker, 2005; Pearson et al., 2002; Usher & Stewart, 2014; D. B. Wilson et al., 2005). Moreover, adherence to the principles of risk, need, and responsivity (RNR)—developed by Andrews and Bonta in the 1990s (Andrews et al., 1990)—in treatment programmes is also associated with reductions in reoffending, although these are based on predominantly quasi-experimental studies (Andrews & Bonta, 2010; Dowden & Andrews, 2000; Hanson et al., 2009). Overall, however, the effectiveness of most prison-based treatments on recidivism remains unclear as the evidence is inconsistent and subject to a range of limitations (Hopkin et al., 2018; Jolliffe & Farrington, 2007; Koehler et al., 2013;

Långström et al., 2013; Lipsey & Cullen, 2007; Lipsey & Wilson, 1998; Papalia et al., 2020, 2019; Perry et al., 2016; Schmucker & Lösel, 2015; Tripodi et al., 2011).

One recent meta-analytic review identified 27 studies of psychological treatments in diverse settings (i.e. prison-based and forensic mental health units) including 7,062 individuals with violent index offences (Papalia et al., 2019). Overall, the results suggested a small, but positive treatment effect on recidivism, with reductions of 11% and 10% for general and violent reoffending, respectively. However, this analysis was largely based on quasi-experimental studies, only half of which (56%) were undertaken in prison environments.

Other previous reviews have often focused on specific groups such as women (Dowden & Andrews, 1999; Tripodi et al., 2011), adolescents (Koehler et al., 2013; Lipsey & Wilson, 1998), individuals who use drugs (Perry et al., 2016), persons living with a mental health condition (Hopkin et al., 2018), and those with sexual (Långström et al., 2013; Schmucker & Lösel, 2015), and violent index offences (Jolliffe & Farrington, 2007). There are considerable methodological differences, particularly in the quality of included primary studies (Koehler et al., 2013), and sources of heterogeneity have rarely been examined (Hopkin et al., 2018). In addition, the vast majority of existing reviews were published before 2008 (Dowden & Andrews, 1999; Jolliffe & Farrington, 2007; Lipsey & Cullen, 2007; Lipsey & Wilson, 1998), and are therefore now outdated. To address these limitations, I aimed to synthesise reoffending outcomes from all RCTs of psychological interventions provided in prisons and jails.

5.2 Methods

5.2.1 Protocol and registration

This systematic review was conducted using PRISMA guidelines (Moher et al., 2009), and the study protocol was preregistered online with PROSPERO (CRD42020167228).

5.2.2 Search strategy

The search strategy was outlined with the assistance of an experienced research librarian from the Bodleian Health Care Libraries (Ms Elinor Harris). I searched the following electronic databases for all relevant surveys from inception until 17 February, 2021: Cochrane Central Register of Controlled Trials, EMBASE, Global Health, MEDLINE, PsycINFO and Google Scholar. The search strategy combined terms relating to randomised controlled trials (i.e., random*, trial*, placebo*), psychological interventions (e.g., program*, intervention*, treatment*), incarceration (e.g., prison*, incarcerat*, custod*) and recidivism (e.g., recommit*, reoffend*, recidiv*) (see Appendix D.1 for a full list of search terms). I also hand-searched reference lists of included studies, relevant articles and systematic reviews.

5.2.3 Study eligibility

I included RCTs of psychological interventions in jails and prisons which reported on criminal recidivism as an outcome. RCTs are the gold standard for investigating therapy efficacy because their rigorous randomisation procedure effectively balances all potential confounders, resulting in experimental and control groups with comparable baseline characteristic profiles. Consequently, it is possible to conclude that the intervention is responsible for any observed differences in clinical outcomes across groups during the follow-up period (Gerstein et al., 2019). Such conclusion cannot be drawn from other types of study designs such as quasi-experimental studies, due to lack of randomisation (Harris et al., 2006).

Studies were eligible for inclusion if they met the following criteria: (1) randomised controlled trials (including pilot studies and cluster-randomised trials); (2) all participants were incarcerated at the time of randomisation (including adolescents, persons on remand, and those living in detention) and remained incarcerated for the duration of the treatment; (3) participants assigned to control groups were exposed to the usual intervention, no intervention or an alternative one; (4) psychological interventions (e.g. CBT or mindfulness-based therapy) or psychoeducational interventions (vocational or educational training); (5) interventions (both individual and group format) were delivered in a jail/prison setting; and (6) outcome of recidivism (e.g. reconviction, reincarceration, rearrest, parole violation or new charges) reported separately for intervention and control groups. We included studies in which post-prison services were offered to participants on a voluntary basis, but were not directly part of the evaluated intervention (e.g. the Challenge to Change (J. Y. Sacks et al., 2012), and the Amity therapeutic community [TC] programmes) (Prendergast et al., 2004).

Studies were excluded on the basis of the following criteria: (1) non-randomised trials (including case studies and pretest/post-test comparisons); (2) participants were not in jail/prison at the time of the study (e.g. on parole, in secure forensic hospitals, attending therapies outside of the prison setting or residing in community-based special residential units (bootcamps); (3) control groups composed of refusals or intervention dropouts; (4) interventions solely based on a pharmacological approach; (5) those that compared jail/prison with a community sanction (i.e. prison vs. boot camp); or (6) joint prison and community programmes, for which the community component accounted for more than half of the duration of intervention (e.g. the CREST programme) (Farrell, 2000; Nielsen et al., 1996). There was no limit on the follow-up time period for reoffending. Non-English language surveys were translated and considered for inclusion.

5.2.4 Data extraction

I performed the searches and screened the titles and abstracts of the surveys identified through the aforementioned search strategy. I also screened the full text of those matching the predetermined inclusion criteria. In cases of uncertainty, I consulted with Dr Rongqin Yu and consensus was reached about study selection. Professor Seena Fazel resolved any disagreements over inclusion and verified the eligibility of included studies.

Information on year of publication, geographical location, correctional setting, sample size, gender, ethnicity (Asian, Black/African American, Caucasian, Hispanic/Latino, Indigenous and Other), average age, follow-up period for recidivism, intervention length, type and format, definition of recidivism and numbers of individuals in the intervention and control groups by recidivism status (i.e. having reoffended vs. having not reoffended) were extracted from eligible studies. In the presence of multiple assessments of recidivism for a given study, I selected the most conservative operationalisation for meta-analysis (e.g. reconviction was preferred to rearrest). As for samples that featured simultaneously men and women and for which the recidivism outcome was not reported separately for each gender, those including $> 90\%$ men were recorded as men, whereas those featuring $< 90\%$ men were recorded as 'both.' When multiple articles were available for a given study (e.g. Amity TC programme) (Wexler, De Leon, et al., 1999; Wexler, Melnick, et al., 1999), we included the one with the longest follow-up period for recidivism (Prendergast et al., 2004). Relevant study authors were contacted when additional data or clarifications were required.

5.2.5 Quality assessment

The quality of randomised studies was assessed using the Cochrane Collaboration's risk-of-bias tool for randomised trials (RoB 2), which is the recommended tool to assess the risk of bias in RCTs (Sterne et al., 2019). This second version supersedes the previous one (RoB), which was initially published in 2008, and subsequently revised in 2011 (Higgins et al., 2011). RoB 2 incorporates existing knowledge of how bias influences study outcomes and how to measure this risk. Each trial was given an overall estimation of risk of bias (i.e. "low risk of bias," "some concerns," or "high risk of bias") according to the following domains: risk of bias arising from the randomisation process, due to deviations from intended interventions, due to missing outcome data, in measurement of the outcome and in selection of the reported result (Sterne et al., 2019). Trials with high risk of bias in any domain(s) were rated as presenting high risk of bias.

5.2.6 Statistical analysis

Effect size calculations

The primary outcome was recidivism. This measure was assessed with the summary odds ratio (OR) and corresponding 95% confidence intervals (CI). In this review, the OR was calculated by dividing the odds that an individual who reoffended had undergone psychological treatment (i.e. was exposed) by the odds that an individual in the control group had reoffended. Both continuous and dichotomous data on recidivism were sought (Webb et al., 2017). To enable comparison across studies, when the outcome was presented as continuous data, I first attempted to obtain the equivalent dichotomous data from the authors of the primary studies. If such information was not available, I converted the standardised mean difference to odds ratios using the following formula according to best practice recommendations (Higgins et al., 2011):

$$\ln\mathrm{OR} = \frac{\pi}{\sqrt{3}}\mathrm{SMD}^{1}$$

One study was excluded due to lack of sufficient information (Kingston et al., 2018). Moreover, for multi-arm trials (Guerra et al., 1990; Shivrattan, 1988), two distinct approaches recommended by the Cochrane Handbook were employed to avoid double-counting participants in the shared control group (Higgins et al., 2011). For one study (Shivrattan, 1988), I merged both intervention arms into one single comparison, as they both were psychoeducational interventions. For another study (Guerra et al., 1990), I included each pairwise comparison separately, as the intervention type differed between them (i.e. one psychoeducational and the other CBT-based), by evenly dividing the shared control group among the comparisons.

Statistical model for meta-analysis

I conducted a random-effects meta-analysis to estimate the effect sizes as it gives similar weights to studies with different sample sizes and significant heterogeneity was expected between studies (e.g. type and length of intervention, follow-up period). Pooled OR estimates were grouped into domains and summarised using forest plots.

[1] Standardised mean difference

Between-study heterogeneity

Between-study heterogeneity was estimated using Cochran's Q (reported with a χ^2-value and p value) and the I^2 statistic. Levels of heterogeneity were evaluated according to previously mentioned thresholds: low (0 to 40%), moderate (30% to 60%), substantial (50% to 90%) and considerable (75% to 100%) (Higgins et al., 2011). These heterogeneity measures should be interpreted with caution when the number of studies is small, particularly in subsequent subgroup analyses (von Hippel, 2015).

Pooling effect sizes

All individual RCTs were pooled to calculate the summary effect size. I then stratified studies according to sample size of psychological-intervention group; ≥ 50 participants. This cut-off was determined in accordance with previous research on randomised experiments (e.g. psychotherapy for adult depression; Cuijpers et al., 2010) to maximise the key beneficial effect of randomisation (i.e. controlling for unknown and unmeasurable variables; Farrington and Welsh, 2005; Welsh et al., 2011), and rule out potential small-study effects (Cuijpers & Cristea, 2016). Amongst these studies, I explored the effects of control group (i.e. usual care, wait-list and other) and intervention type (i.e. CBT-based, psychoeducational, therapeutic communities and other) with the removal of two studies (Annis, 1979; Shapland et al., 2008). I excluded them from this secondary analysis on the basis of considerable differences in treatment duration (e.g. one session only; Shapland et al., 2008) and delivery mode (e.g. video feedback of previous sessions; Annis, 1979).

All interventions which were based on cognitive behavioural approaches were considered to be CBT-based psychological interventions. Those that present a core vocational or educational component, such as deterrence (Lewis, 1983), were included in the psychoeducational category (Lattimore et al., 1990). Interventions of therapeutic communities formed another category (Prendergast et al., 2004; J. Y. Sacks et al., 2012). Both TC trials included voluntary post-prison services. Most participants from the Challenge to Change trial (83%) chose to access community-based mental health or substance abuse services, although these were beyond the scope of the study (J. Y. Sacks et al., 2012). The Amity TC offered residential treatment to program graduates (experimental group only) at an Amity-operated facility (i.e. Vista; Prendergast et al., 2004). The impact of Vista on recidivism was not considered in this meta-analysis to avoid annulling the effects of randomisation. However, I report percentages in the Discussion of this chapter. The other category combined reality therapy (Dugan & Everett, 1998), social therapy (Ortmann, 2000), interactive journaling (Proctor et al., 2012), and gender-responsive substance abuse therapy (Messina et al., 2010).

Subgroup analyses and meta-regression

Pre-specified subgroup and meta-regression analyses were conducted to examine sources of heterogeneity. I implemented a mixed-effect model for the subgroup analyses, which assumes a random-effects model within subgroups and a fixed-effect model between subgroups. It should be noted that this method presupposes homoscedasticity, which holds that residual heterogeneity is the same across subgroups (Borenstein & Higgins, 2013; Harrer et al., 2021). The following study characteristics were assessed: year of publication ($<$1990 vs. $\geq$1990; to account for the formalisation of the RNR model in 1990; Andrews et al., 1990), study origin (US vs. elsewhere), sample size (as a continuous variable), sex (sex-specific interventions vs. those being delivered to both men and women simultaneously), mean age (as a continuous variable), age group (adolescents vs. adults), intervention type (CBT-based vs. all other types), comparator type (usual care vs. waitlist/other), follow-up time period (as a continuous variable), intervention format (individual vs. group/combination), intervention aimed at SUD (as a dichotomous variable) and risk of bias (high vs. low/some concerns).

Influence analysis

I performed influence analysis on all studies to determine which of them influenced disproportionately the summary effect of mymeta-analysis. I employed the Leave-One-Out method and presented results using the Baujat plot. The latter is a graphical method to identify studies that contribute to overall heterogeneity in a significant manner. Specifically, this diagnostic plot illustrates the heterogeneity contribution of each RCT (on the horizontal axis) in relation to their individual influence on the pooled effect size estimate (on the vertical axis) (Baujat et al., 2002).

I examined publication bias in all studies using the Egger's regression test (Egger et al., 1997) and funnel plot analysis. If the Egger's test reported publication bias and between-study heterogeneity was not substantial (Peters et al., 2007), I followed the 'trim-and-fill' procedure (Duval & Tweedie, 2000). This funnel-plot derived approach is intended to identify and correct publication bias-related funnel plot asymmetry (Higgins et al., 2011; Mavridis & Salanti, 2014). To do so, studies that contribute to the funnel plot's asymmetry are first trimmed, following which a new overall effect estimate is calculated based on the remaining studies (i.e. those that can be considered minimally impacted by publication bias). Finally, the 'missing' studies are imputed into a new symmetrical funnel plot using the bias-corrected overall estimate (Schwarzer et al., 2017; Shi & Lin, 2019).

Sensitivity analysis

In the presence of small-study effects indicated by the results of the publication bias analysis, I conducted sensitivity analyses. First, the fixed-effect and random-effects estimates of the intervention effect were compared, as a more favourable estimate in the random-effects model might indicate that interventions were more effective in smaller studies. This can be explained by the fact that random-effects summaries are more sensitive to publication bias than fixed-effect summaries (Poole & Greenland, 1999). I conducted an additional analysis by only including studies with an intervention group of $\geq$100 participants. This was done to mitigate small-study effects, and evaluate the robustness of the findings, as small trials are vulnerable to selection bias and tend to have larger and more favourable treatment effects than large trials (Schwarzer et al., 2017; Sterne & Egger, 2001). I also investigated the impact of study quality on the pooled effect size by removing studies at high risk of bias (based on the risk assessment conducted with ROB 2). All statistical analyses were performed in R/R Studio (R Core Team, 2021; RStudio Team, 2021) following guidance from Harrer et al. (2021)'s *Doing Meta-Analysis with R* guide (online version). P values less than .05 were considered statistically significant.

5.3 Results

I identified 6,345 articles through electronic searches and 29 eligible trials (k = 29; see Figure 5.1 for selection process and Table 5.1 for study characteristics). Most RCTs were two-arm trials (k = 27); two were three-arm trials (k = 2). These trials described 31 psychological interventions which were combined into 30 pairwise treatment comparisons (n = 30), and on which the statistical analyses were based. In total, 9,443 individuals (1,104 women [11.7%], 8,111 men and 228 individuals where sex was not reported) participated in the trials, and 6,528 (1,118 [17.1%] adolescents and 5,410 adults) had recidivism outcome data. The mean age was 31.4 years (SD = 4.9; range 24.5–41.5) in adults and 17.5 (SD = 1.9; range 14.6–20.2) in adolescents. Descriptive statistics on the age of participants were calculated using the mean age from each study and the range of mean ages (where available). Ethnicity data from each study is summarised in Appendix D (see Table D.2). Amongst included trials, 19 were from the US (n = 3,578 [54.8%]); four were from Canada (n = 2,351); two were from the United Kingdom (n = 203); and one each were from Germany (n = 223), Sweden (n = 59), Japan (n = 50), and Norway (n = 64). Treatment duration considerably varied between trials, ranging from one session only to multiple interventions that lasted for one year. The most frequent source of funding for trials was government-funded research councils. None of the psychological interventions were described as being mandatory and recruitment of participants was voluntary. However, it is possible that perceived coercion and other incentives could have contributed to the decision to participate.

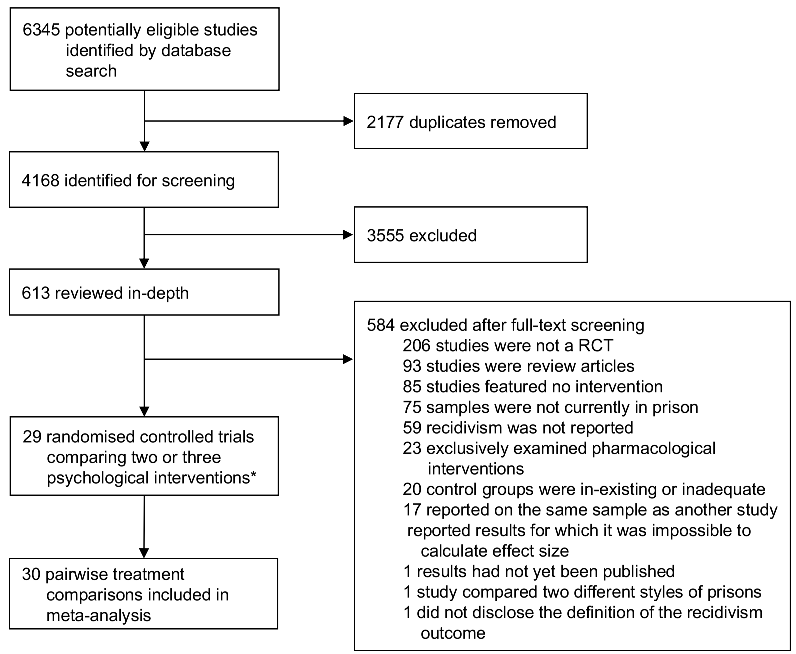


Figure 5.1: Study selection. *The 29 randomised controlled trials, included 27 RCTs that were two-arm trials and two that were three-arm trials (Guerra et al., 1990; Shivrattan, 1988). Overall, the trials described 31 psychological interventions that were combined into 30 pairwise treatment comparisons on which the statistical analyses were based.

Table 5.1: Characteristics of randomised controlled trials of psychological interventions

Study	Country	Setting	Number of randomised participants	Number of participants followed-up	Sex	Mean age	Psychological intervention, type and format	Comparator type	Duration of the intervention and number/frequency of sessions	Detailed definition of recidivism outcome	Follow-up period of recidivism
Persons, 1967	USA	Institution for boys	82	82	Men	16.4	Psychotherapy; other; combination	No treatment	20 weeks (80h and 60 sessions) total; twice weekly for group psychotherapy (1.5h per session) and an average of 1h a week of individual psychotherapy	Reinstitutionalisation in any penal institution	$M = 9.5$ months
Annis, 1979	Canada	Minimum-security institution	150	128	Men	24.5 (range 18–64)	Awareness group (with and without video feedback); psychoeducational; group	Routine institutional care	8 weeks total; $M = 224$ program hours	Incarcerated at follow-up	1 year
Lewis, 1983	USA	Four camps	108	108	Men	16.3 (range 14–18)	Squires program; psychoeducational; group	No treatment	3 consecutive mornings; 3 hour sessions	Subsequent arrests and/or charges	1 year
Linden and Perry, 1984	Canada	Two penitentiaries (maximum- and medium-security)	66	55	Men	Not reported	Prison education program; psychoeducational; combination	No treatment	Not reported	Marginal failure (return to prison for minor crime or technical violation of parole regulations) and clear recidivist (return to prison for major offense)	77–82 months

Table 5.1: Characteristics of randomised controlled trials of psychological interventions (continued)

Study	Country	Setting	Number of randomised partici-pants	Number of partici-pants followed-up	Sex	Mean age	Psychological interven-tion, type and format	Comparator type	Duration of the inter-vention and num-ber/frequency of sessions	Detailed definition of recidivism outcome	Follow-up period of recidivism
Homant, 1986	USA	Prison	92	86	Men	Not reported	Group therapy; other; group	Standard care (control subjects were free to seek out therapy [group or individual] through normal channels)	M number of therapy contacts during the first year of institu-tion = 18.6 (experimen-tal group) and 4.0 (control group)	Reincarceration on felony (currently under supervision) and currently reincarcer-ated for a new felony	10 years
Shivrattan, 1988	Canada	Institution for incarcerated delinquents	45	42	Men	Range 15–17	Social interaction skills program and stress management training program; psychoeduca-tional; group	No treatment	8 sessions	Further criminal activity (being charged and sentenced to an institution)	12 to 15 months
Guerra et al., 1990	USA	Juvenile correctional facility	165	83	Both (women [50%] and men [50%])	17.2 (range 15–18)	CMT and AC; CBT-based and psychoe-ducational; group	No treatment	12 weeks total; 12 sessions	Parole violation	At least 1 year and up to 2 years
Lattimore et al., 1990	USA	Prison	591	247	Men	20.0	Vocation delivery system; psychoeduca-tional; group	Routine care (e.g. assignment to the first available vocational training or to a prison job)	Not reported	Rearrest	M = 2 years (range 411 to 1,530 days)
Leeman et al., 1993	USA	Medium-security correctional facility	57	57	Men	16.0	EQUIP; CBT-based; group	Simple or motivational	1 to 1.5 hours, 5 days per week	Parole revocation and/or institutional recommit-ment	6 and 12 months
Robinson, 1995	Canada	Correctional facility	4,072	2,125	Men	29.6	Cognitive skills training; CBT-based; group	Waitlist	36 sessions	Reconviction for a new offence	1 year

Table 5.1: Characteristics of randomised controlled trials of psychological interventions (continued)

Study	**Country**	**Setting**	**Number of randomised partici-pants**	**Number of partici-pants followed-up**	**Sex**	**Mean age**	**Psychological interven-tion, type and format**	**Comparator type**	**Duration of the inter-vention and num-ber/frequency of sessions**	**Detailed definition of recidivism outcome**	**Follow-up period of recidivism**
Lindforss and Magnusson, 1997	Sweden	Prison	60	59	Men	Not reported	SFBT; other; individual	No treatment	Not reported	Further offense which has resulted in a sentence to probation or imprison-ment	12 and 16 months
Dugan and Everett, 1998	USA	Jail	145	117	Men	30.2 (SD = 9.0)	Reality therapy; other; group	No treatment	72 hours total	Mean number of offense charges	2 years
Ortmann, 2000	Germany	Prison	228	223	Not reported	Not reported	Social therapy; other; not reported	No treatment	Not reported	All new sentences	5 years
Armstrong, 2003	USA	YOU in a detention center	256	212	Men	20.2 (range 15–22; SD = 1.0)	MRT; CBT-based, group	No treatment	1 to 1.5 hours, on average 3 sessions per week	Arrest followed by a conviction for which jail or prison time was levied and served	M = 563 days (treatment group) and 617 days (control group)
Prendergast et al., 2004	USA	Medium-security prison	715	576	Men	30.9	Amity therapeutic community program; therapeutic communities; group	No treatment	1 year total	Reincarceration	5 years
S. Sacks et al., 2004	USA	Prison	236	107	Men	34.3 (SD = 8.8)	Prison MTC + aftercare; therapeutic communities; group	Mental health treatment program	1 year total	Reincarceration	1 year
Shapland et al., 2008	UK	Prison	94	94	Men	JRC Restorative justice scheme; other; individual	No treatment	One conferencing session	Reconviction	2 years	

Table 5.1: Characteristics of randomised controlled trials of psychological interventions (continued)

Study	Country	Setting	Number of randomised participants	Number of participants followed-up	Sex	Mean age	Psychological intervention, type and format	Comparator type	Duration of the intervention and number/frequency of sessions	Detailed definition of recidivism outcome	Follow-up period of recidivism
Zlotnick et al., 2009	USA	Residential substance abuse treatment program in a minimum security wing of a women's prison	49	44	Women	34.6 (SD = 7.4)	Seeking Safety + TAU; CBT-based, group TAU (similar to other state prison substance use program)	6–8 weeks total; 90 min, typically 3 times per week	Reincarceration	6 months	
Messina et al., 2010	USA	Prison for women	115	115	Women	35.9 (SD = 9.6)	GRT using manualized curricula (Helping Women Recover and Beyond Trauma); other; group	Standard prison therapeutic community program	Helping Women Recover (17 sessions) and Beyond Trauma (11 sessions)	Reincarceration	1 year
Proctor et al., 2012	USA	Jail	185	183	Men	36.6 (SD = 11.1)	Interactive journaling; other; individual	Placebo (government booklet on substance abuse disorders and criminal behavior)	Not reported	Being booked in the county jail	1 year
J. Y. Sacks et al., 2012	USA	Women's correctional facility	468	370	Women	35.1 (SD = 7.9)	Challenge to change therapeutic community; therapeutic communities; group	Cognitive behavioral substance abuse intervention	Planned 6-month tenure; program activities were provided 4h per day, 5 days per week	Reincarceration	1 year
Bowes et al., 2014	UK	Two medium-security prisons	115	109	Men	24.5 (SD = 5.7)	COVAID + TAU; CBT-based; group	TAU	4 weeks total; 10 sessions; approx. 20h of group treatment and at least 4h of individual support	Reconviction	$M = 518$ days

Table 5.1: Characteristics of randomised controlled trials of psychological interventions (continued)

Study	Country	Setting	Number of randomised participants	Number of participants followed-up	Sex	Mean age	Psychological intervention, type and format	Comparator type	Duration of the intervention and number/frequency of sessions	Detailed definition of recidivism outcome	Follow-up period of recidivism
Yokotani and Tamura, 2015	Japan	Prison	50	50	Men	41.5 (SD = 10.5)	PFI; other; individual	No treatment	Letters were sent over the course of 3 months; 6 personalised feedback letters; 2 times per month	Reentering prison	M = 3.6 (range, 0.1–5.8 years)
Chaple et al., 2016	USA	Ten prisons	494	482	Both (women [31.4%] and men [69.6%])	36.6 (SD = 9.6)	E-TES; CBT-based; individual	Standard care	12 weeks total; 48 interactive, multimedia modules; once a week for 2h or twice weekly for 1h (depending on lab availability)	Reincarceration	1 year
Kubiak et al., 2016	USA	Prison for women	42	35	Women	33.7 (SD = 8.9)	Beyond violence; other; group	TAU	20 sessions; 40 hours total	Reincarceration	1 year
Burraston and Eddy, 2017	USA	Four state correctional facilities (releasing institutions)	359	359	Both (women [55%] and men [45%])	31.4 PMT; CBT-based; group	Services as usual	12 weeks total; 2.5h sessions, 3 times per week	Mean (count) of postrelease arrests	1 year	
Malouf et al., 2017	USA	Jail	49	31	Men	37.2 (range, 18–81)	REVAMP + TAU; other; group	TAU	4 weeks total; 90-minute sessions delivered twice per week	Rearrest	3 years
Hein et al., 2020	USA	Juvenile justice setting	289	289	Men	14.9 (SD = 1.0)	SPST; CBT-based; group	TAU	10 one-hour sessions	At least one offense during follow-up period	2 years
Gold et al., 2021	Norway	Prison	66	64	Men	26 (range, 18–53)	Music therapy; other; usually group, but in some cases individual	Standard care	Mean = 4.4 (0–12); Typically 2–3 /week)	Serious events, excluding writs	5 years

Note. AC = attention control; CBT = cognitive behavioural therapy; CMT = cognition mediation training; COVAID = control of violence for angry, impulsive drinkers; EQUIP = equipping youth to help one another; E-TES = experimental condition, therapeutic education system; GRT = gender responsive therapy; JRC = justice research consortium; MRT = moral reconation therapy; MTC = modified therapeutic community; PFI = personalised feedback intervention; PMT = parent management training; REVAMP = re-entry values and mindfulness program; SFBT = solution focused brief therapy; SPST = social problem solving training; TAU = treatment as usual; YOU = young offenders unit.

5.3.1 Risk of bias of the included studies

Most trials were rated as presenting with concerns of bias (60%) or being at high risk of bias (33%). Only two (Burraston & Eddy, 2017; Messina et al., 2010) were rated as low risk of bias (Appendix D Table D.2). There was a low risk of bias in outcome measurement for all as recidivism was ascertained from official criminal records.

5.3.2 Meta-analysis

Overall, psychological interventions were associated with reduced reoffending, with a pooled OR of 0.72 (95% CI 0.56–0.92) and moderate levels of heterogeneity (I^2 = 49%; Q = 57.3; $p < .01$, Figure 5.2). To prevent against overestimation caused by small-study effects suggested by the literature and confirmed by the influence analysis, I pooled results excluding studies with less than 50 participants in the experimental arm. The reduction in recidivism was attenuated in trials with an intervention group of $\geq$ 50 participants (OR = 0.87, 95% CI 0.68–1.11; I^2 = 54%; Figure 5.3).

Subgroup analyses are presented in Figures 5.4 (by comparator type) and 5.5 (by intervention type). RCTs with control group of usual care were associated with recidivism but not significantly (OR = 0.97, 95% CI 0.70–1.34; I^2 = 59%). When using waiting list (OR = 0.74, 95% CI 0.56–0.99; I^2 = 17%) or other interventions (OR = 0.64, 95% CI 0.40–1.01; I^2 = 0%), the reduction in recidivism was stronger although CIs were overlapping. By treatment modality, CBT-based interventions were not associated with recidivism (OR = 1.00, 95% CI 0.69–1.44; I^2 = 60%) neither were psychoeducational interventions (OR = 1.11, 95% CI 0.38–3.20; I^2 = 79%). Other types of interventions were associated with non-significant reductions in recidivism (OR = 0.74, 95% CI 0.47–1.18; I^2 = 44%). However, there were reductions in reoffending risk for TC programmes (OR = 0.64, 95% CI 0.46–0.91; I^2 = 0%).

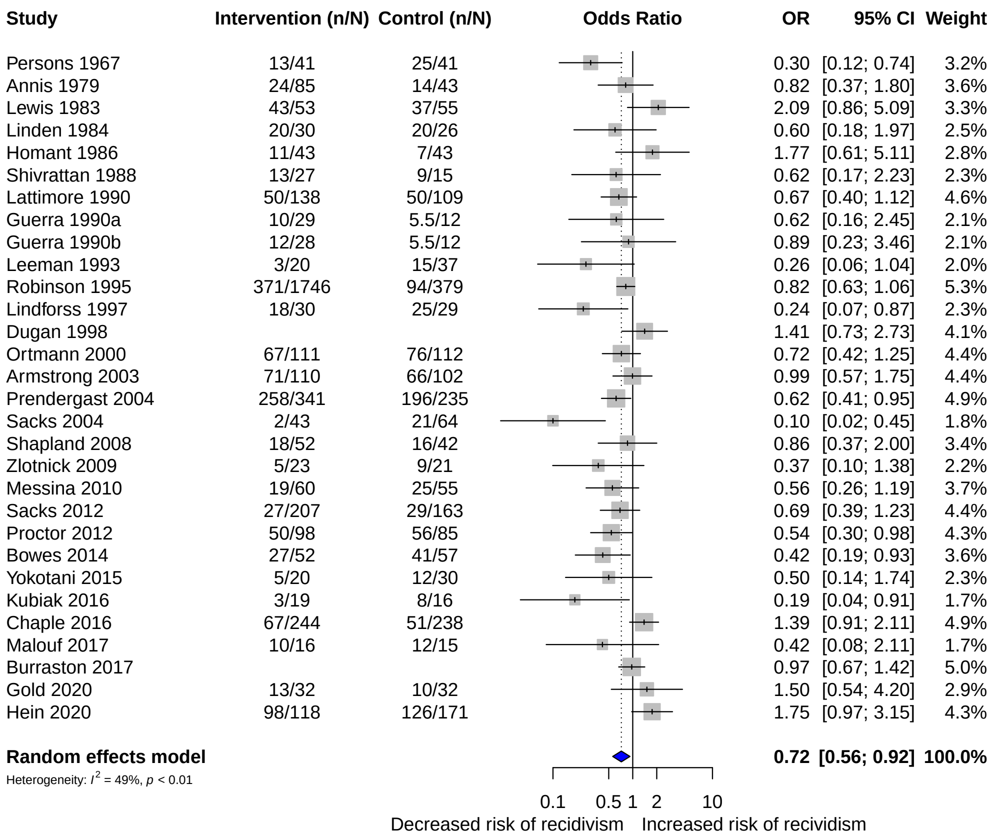


Figure 5.2: Forest plot of all studies (random-effects model). Numbers of participants in the intervention and control groups are not available for Dugan and Everett (1998) and Burraston and Eddy (2017), as the outcome was presented as continuous data rather than dichotomous data in both of these studies.

5.3.3 Heterogeneity analysis

On univariate analyses, there was a statistically significant difference between the pooled effects of trials which included gender-specific samples when compared to those which did not ($Q = 4.30$, $p = 0.04$). Gender-specific RCTs had associations with reduced recidivism (OR = 0.67; 95% CI 0.50–0.90), whilst those including both men and women did not (OR = 1.09; 95% CI 0.77–1.55). No other significant associations were found between effect sizes and pre-specified study characteristics in subgroup or meta-regression analyses (Table 5.2).

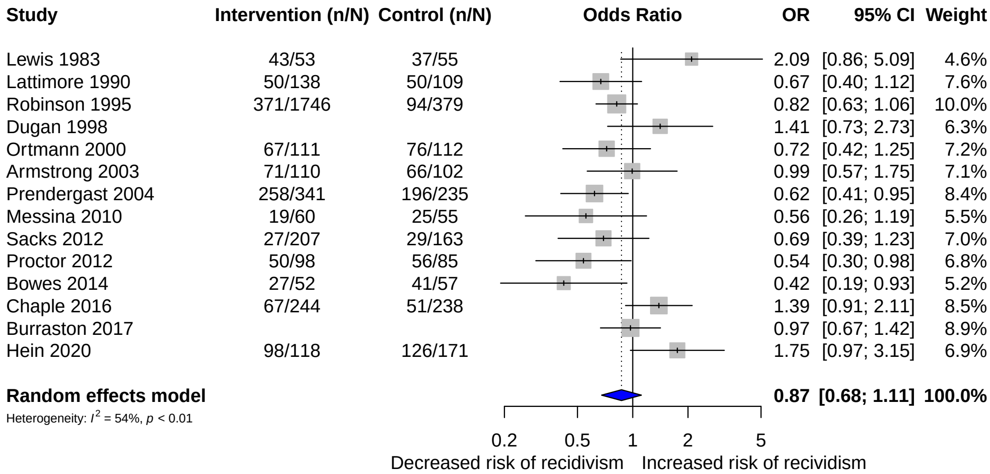


Figure 5.3: Effectiveness of psychological interventions in prison in reducing recidivism. Data are reported on studies with an intervention group of ≥50 participants, excluding two outliers (Annis, 1979; Shapland et al., 2008). Error bars represent odds ratios 95% confidence intervals (CI).

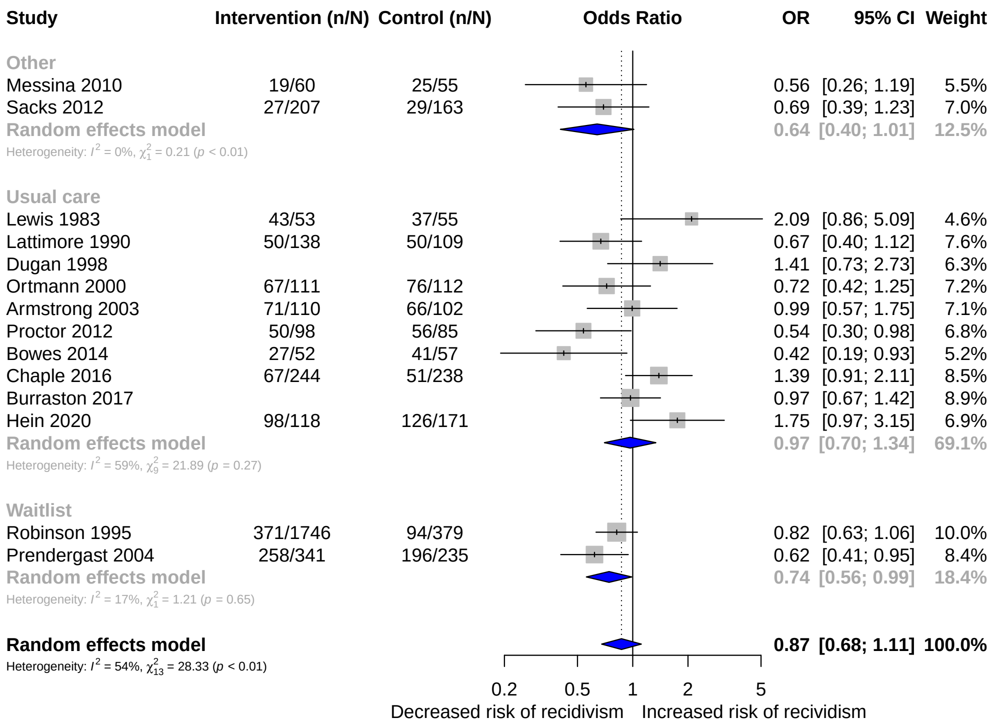


Figure 5.4: Effectiveness of psychological interventions in prison in reducing recidivism by comparator type. Data are reported on studies with an intervention group of ≥50 participants, excluding two outliers (Annis, 1979; Shapland et al., 2008). Error bars represent odds ratios 95% confidence intervals (CI).

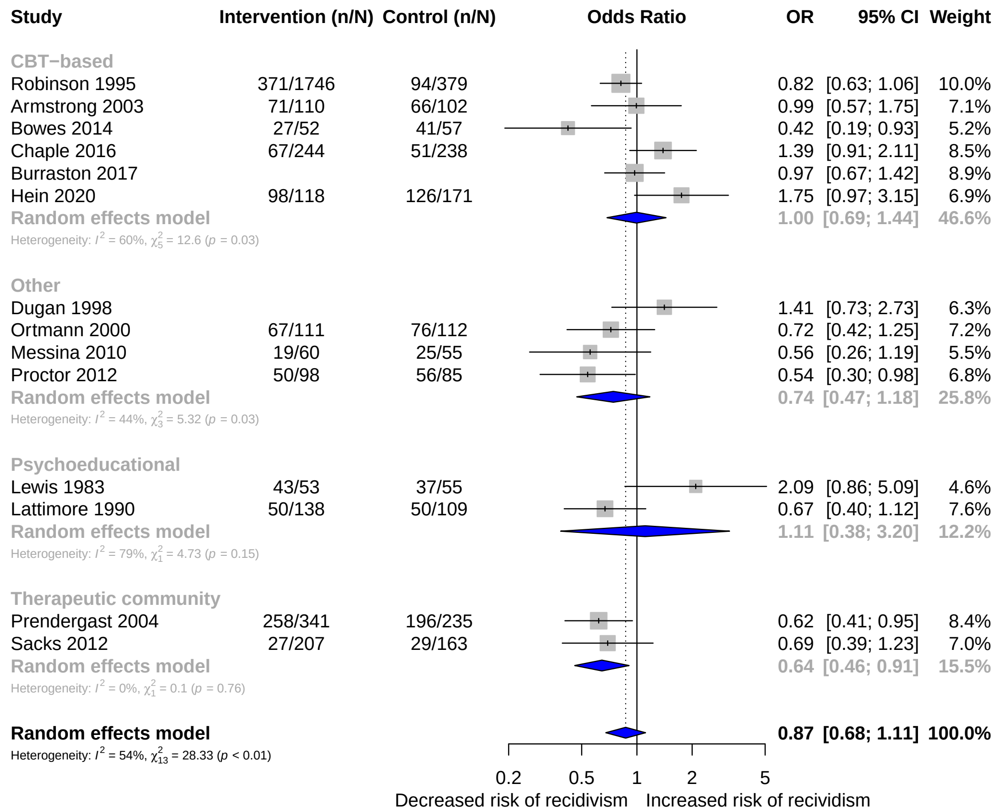


Figure 5.5: Effectiveness of psychological interventions in prison in reducing recidivism by intervention type. Data are reported on studies with an intervention group of ≥50 participants, excluding two outliers (Annis, 1979; Shapland et al., 2008). Error bars represent odds ratios 95% confidence intervals (CI). CBT = cognitive behavioural therapy.

5.3.4 Influence analysis

Two studies (Hein et al., 2020; S. Sacks et al., 2004) that contributed disproportionately to the pooled effect were identified using influence analyses in all RCTs. Removal of these outliers reduced the degree of heterogeneity between studies from moderate (I^2 = 49%) to low (38%), but did not materially alter the pooled effect size (OR 0.73; 95% CI 0.58–0.91; Appendix D Figures D.2—D.4).

I found evidence of publication bias using Egger's test (t = -2.12, p = 0.04) suggesting small-study effects. This finding was supported by visual inspection of

Table 5.2: Meta-regression analyses assessing the link between study characteristics and risk of recidivism

Variable	β	**SE**	p
Year of publication: ≥1990 vs. <1990	-.195	.335	.560
Study origin: USA vs. elsewhere	.097	.274	.722
Sample size (continuous)	.000	.000	.671
Sex: sex-specific interventions vs. both	-.404	.371	.276
Mean age (continuous)	-.016	.018	.372
Age group: adolescents vs. adults	-.161	.284	.570
Intervention type: CBT-based vs all other types	-.217	.270	.422
Comparator type: usual care vs. waitlist/other	.396	.301	.189
Follow-up time period (continuous)	.074	.063	.239
Intervention format: individual vs. group/combination	-.055	.348	.875
Intervention aimed at PIP with a SUD (dichotomous)	-.283	.256	.269
Risk of bias: high vs. low/unclear	-.146	.266	.583

Note. CBT = cognitive behavioural therapy; PIP = people in prison; SUD = substance use disorder.

the related funnel plot, which showed asymmetry (Appendix D Figures D.5 and D.6). Seven smaller studies were identified and trimmed using the trim-and-fill method (Kubiak et al., 2016; Leeman et al., 1993; Lindforss & Magnusson, 1997; Persons, 1967; S. Sacks et al., 2004; Shivrattan, 1988; Zlotnick et al., 2009), and the OR adjusting for publication bias was 0.86 (95% CI 0.65–1.15).

5.3.5 Sensitivity analysis

The fixed-effect estimate (OR = 0.81; 95% CI 0.72–0.91; I^2 = 49%; Appendix D Figure D.7) did not materially differ from the random-effects model. Repeating the meta-analysis by only including larger studies (i.e. ≥100 participants in the psychological-intervention group) resulted in a decrease of the strength of the association to an OR of 0.90 (95% CI 0.71–1.14; Appendix D Figure D.7; Cuijpers et al., 2010).

5.4 Discussion

This chapter provides new evidence on the effectiveness of psychological therapy for reducing recidivism, by being the first meta-analysis of RCTs, to my knowledge, to study these treatments in jails and prisons. The effects are considerably smaller than expert opinion has previously maintained based on 29 jail/prison–based RCTs of 9,443 individuals from 7 countries. I conducted rigorous sensitivity, heterogeneity and influence analyses to evaluate the robustness of effect estimates. Thus, studies with less than 50 people in experimental arms were excluded to account for small-study effects, resulting in 14 trials with 6,446 followed-up participants. I observed an overall pooled OR of 0.87 (95% CI 0.68–1.11), indicating at most modest effects.

5.4.1 Main findings

This investigation highlights two other main findings. First, I found no strong evidence of reduced reoffending following participation in CBT-based prison-based programmes (OR = 1.00, 95% CI 0.69–1.44; $I^2 = 60\%$). This is in contrast with a 2007 systematic review combining both prison and community based interventions that reported reduced risks of 20-30% (Lipsey & Cullen, 2007). One potential explanation for the lack of clear effectiveness of such CBT interventions is that they are not linked with psychosocial support upon release. It may also be that these psychological therapies, which were developed for mental health problems, do not address accommodation, employment, and financial difficulties after release that contribute to recidivism risk (Hirschtritt & Binder, 2017).

A second finding from the chapter is that participation in a therapeutic community was associated with reduced reoffending risk. However, this finding was limited to only two studies (Prendergast et al., 2004; J. Y. Sacks et al., 2012), both of which linked released prisoners to voluntary post-prison services. In support, in one of these trials, links to community services were associated with a lower return to custody rate than those who did not have these links (42% return rate vs. 86%; Prendergast et al., 2004). Findings from a recent systematic review of psychoeducational programmes for reducing prison violence are consistent with the potential role of TCs, as those tailored to specific needs (e.g. substance use disorder) were associated with reduced institutional violence (Auty et al., 2017). Similar results were reported in a Cochrane review of any persons who offended and had co-occurring drug and mental health problems (Perry et al., 2019a), as three (J. Y. Sacks et al., 2012; S. Sacks et al., 2004; Wexler, De Leon, et al., 1999) of the four included studies (J. Y. Sacks et al., 2012; S. Sacks et al., 2004; Shapland et al., 2008; Wexler, De Leon, et al., 1999) found TCs were associated with reductions in recidivism.

5.4.2 Implications

There are several implications of this chapter for treatments offered in prison. First, in-prison interventions may not be effective if they are not linked with interventions that target psychosocial needs of released individuals. As such, two TC trials highlighted the potential importance of community aftercare to maintain the therapeutic gains delivered in prison (Prendergast et al., 2004; J. Y. Sacks et al., 2012). Hence, psychological interventions which combine prison and community-based services should be prioritised for future research. As part of this, it should be noted that recent efforts to implement 'through the gate' care in the UK have been widely criticised due to lack of communication between prisons and community services, and poor screening of people in prison early in their sentences for resettlement needs (HM Inspectorate of Probation and HM Inspectorate of Prisons, 2016). There are also systemic barriers to widespread uptake in the US, where individuals with a previous conviction have limited access to public housing assistance, health and social services, and few employment opportunities (Raphael, 2011).

Second, most of the tested interventions were developed in the community or clinical populations for other outcomes, and hence they may not address risk factors specific for reoffending. Any such risk factors may need to be identified by high quality assessment, which can then be linked to interventions for reducing recidivism. As such, risk assessments should be informed by scalable and transparent clinical prediction tools, such as OxRec which includes assessment of modifiable risk factors for recidivism (e.g. substance misuse and mental health status; Fazel, Chang, et al., 2016), supplemented by more detailed assessments that consider additional dynamic factors. Considering that resources allocated for interventions in prison populations are limited (National Research Council, 2008), stratification of risk is necessary to guide risk management and treatment of people on release from prison.

A third implication is regarding CBT. The modest effects that we reported are different to evidence from some reviews, including one published by the Campbell Collaboration (Lipsey et al., 2007), which have suggested that CBT is one of the most effective form of treatment for people in prison (Henwood et al., 2015; Landenberger & Lipsey, 2005; Lösel & Schmucker, 2005; Pearson et al., 2002; Usher & Stewart, 2014; D. B. Wilson et al., 2005). However, these previous reviews combined RCTs with less rigorous study designs including matched-subjects design. In addition, the findings question the widespread roll-out of these treatment approaches in prisons. In fact, only one (Bowes et al., 2014) of the six CBT studies (Armstrong, 2003; Bowes et al., 2014; Burraston & Eddy, 2017; Chaple et al., 2016; Hein et al., 2020; Robinson, 1995) in the current review reported significant reductions in reoffending. Other research in more selected populations of all people who have offended and use drugs has also found little support for CBT (Perry et al., 2019a; Perry et al., 2019b).

Another implication of the current chapter is that the effects of in-prison psychological interventions on recidivism appear to be smaller than those reported in previous meta-analyses, which have been estimated as around 0.65 (95% CI 0.57–0.75; Papalia et al., 2019). This is likely due to previous reviews including poorer research designs, such as quasi-experimental studies (Weisburd et al., 2001). This has also been found for psychotherapy effectiveness in depression, where overall effectiveness was overestimated in earlier meta-analyses due to inclusion of non-experimental designs (Cuijpers et al., 2010).

5.4.3 Research gaps

My analyses also highlight several evidence gaps. More research is required to determine whether generic psychological interventions are effective in specific groups of incarcerated populations, such as those living with mental disorders other than substance misuse. Research to date suggests that tailored individualised interventions are associated with better treatment outcomes (Fontanarosa et al., 2013). Moreover, to improve transition to the community, future research should develop and evaluate the impact of follow-up community treatments in the community after release. Greater consideration should be given to understanding the influence of environmental factors within prisons on treatment effects. Potential effects could be limited due to prisons not being primarily therapeutic environments, and prioritising security over health and rehabilitation needs (Yoon et al., 2017). To understand this, research comparing the effectiveness of the same treatment modality in prison with community settings can provide information on how the prison environment affects sustained behavioural change and what adaptations could improve their effectiveness in prisons. Technology-based psychological interventions could also potentially be leveraged to mitigate limited resources in custodial settings and in turn improve reoffending outcomes for people who experience incarceration, although these were beyond the scope of this chapter. Some research has shown promising effects (e.g. Grove et al., 2021), but RCTs of virtual therapy within prisons are still few and far between.

5.4.4 Limitations

To the best of my knowledge, this chapter describes the first meta-analysis of high-quality RCTs of the effectiveness of psychological interventions; however, there are several limitations. The study selection process leading up to the full-text screening stage was conducted by a single reviewer. The included trials were delivered in HICs. The reason for this could be two-fold: either very few to no psychological interventions are implemented in justice settings in LMICs, or some do exist, but they are not researched in current practice (particularly when it comes to experimental study designs). Another review expanded the search to include quasi-experimental designs, yet still no studies from LMICs were found (Papalia et al., 2019). In addition, the number of included studies was not large (k = 29), which underlines the legal, practical and ethical challenges of conducting high-quality research in prisons (Farrington & Welsh, 2005; Quina et al., 2007; Yoon et al., 2017). One specific problem encountered when conducting clinical research in these settings is high dropout rates, which often result in small and selected samples. Prisons and especially jails have high turnover rates and, participants are likely to be released or transferred unexpectedly (Lobmaier et al., 2010). There were not sufficient studies to produce meta-analysis results for violent reoffending—notwithstanding it being the main focus of this thesis—and thereby findings from general recidivism need to be extrapolated until further research is carried out.

Furthermore, despite limiting inclusion to the most robust study design of RCTs, only 7% of the included studies had low risk of bias. The most affected domains were randomisation and deviations from the intended interventions. The difficulty associated with blinding staff and participants to the assigned intervention is likely to have contributed to increased risk of bias in these two domains. There was also evidence of selective publication of smaller studies based on their effect size (i.e. some of those with small effect sizes were missing), which indicated that myinitial pooled estimate of all studies (OR = 0.73) was overestimated due to publication bias (Borenstein & Higgins, 2013). Sex-specific analyses comparing estimates in women and men could not be performed due to the insufficient numbers of women-only samples.

5.5 Conclusion

5.5.1 What this chapter adds

This chapter provides a synthesis of current research on the effectiveness of psychological interventions delivered in prisons aimed at reducing post-release recidivism. The findings differ from those of previous meta-analyses, and the effectiveness of psychological interventions delivered in prison is therefore less than previously assumed, with at best modest effects. Most intervention effects, particularly for CBT, are small and statistically insignificant, but some merit additional study and replication. Consequently, trials of TC interventions and related approaches that facilitate the continuity of treatment after prison release should be prioritised. Considering high rates of recidivism (Petersilia, 2011; Yukhnenko et al., 2019), and its consequences for public health and safety (Newton et al., 2019; Steinfeld et al., 2018), simple large RCTs on the effectiveness of psychological interventions in prison are necessary.

Contribution statement

This chapter is based on the published work: *Beaudry, G., Yu, R., Perry, A. E., & Fazel, S. (2021). Effectiveness of psychological interventions in prison to reduce recidivism: a systematic review and meta-analysis of randomised controlled trials. Lancet Psychiatry, 8(9), 759–773.* I designed the study, conducted the data search, extraction, analyses, and prepared the tables and figures under the supervision of Prof Seena Fazel and Dr Rongqin Yu. I also wrote the first draft of the manuscript and implemented the contribution of the co-authors and external reviewers up to final publication.

5.5.2 Next chapter

Together, the findings of the present chapter indicate that there is only minimal evidence to support the effectiveness of widely implemented psychological interventions for people in prison to reduce reoffending; thus, answering the third thesis research question (see Chapter 1). The next chapter (Chapter 6) will provide a general discussion of the thesis findings, and their implications for clinical practice, research and policy. I will also discuss potential new avenues of research that could build on my doctoral work, and more broadly on recent research in risk prediction and prison public health. I will then present overarching strengths and limitations, and finally provide concluding remarks for this programme of research.

6
General discussion

Contents

6.1 Summary of the thesis

This thesis aimed to provide a comprehensive body of research relating to violent reoffending amongst people released from prison, with a particular focus on violence risk assessment, modifiable risk factors and psychological interventions. I employed an evidence synthesis approach (systematic review and meta-analysis) to answer research questions (**R1** & **R3**) by identifying and summarising all the relevant individual prevalence and interventions studies, as relevant. Advanced statistical methods were also used to evaluate the predictive validity of an existing violence risk prediction model (OxRec) in new samples of individuals who have experienced incarceration (**R2**). In this final chapter, I will specifically address the research questions previously detailed in Chapter 1 by summarising these analytical procedures and the principal findings from the aforementioned empirical studies. Further, I will discuss the originality of my doctoral research programme, as well as its general strengths and limitations. Implications for research, policy and practice will be detailed. I will also explore remaining knowledge gaps and potential future directions as informed by the current scientific literature.

6.2 Overview of findings

6.2.1 What is the prevalence of mental disorders amongst adolescents in juvenile detention and correctional facilities?

Chapter 2 synthesised the global prevalence of treatable mental disorders amongst adolescents in juvenile detention and correctional facilities. Building on a previous meta-analysis of 25 surveys of psychiatric morbidity (Fazel, Doll, et al., 2008), I broadened the scope of search by including an additional disorder (PTSD) and extending the time period of included studies by more than ten years. This updated search strategy identified 22 new prevalence studies, bringing the total to 47 studies from 19 different countries, none of which were LMICs. Major depression (about 1 in 10) and ADHD (about 1 in 5) were the two most prevalent mental disorders in male adolescents. One in four female adolescents were also diagnosed with major depression, and one in five had current PTSD. I found that major depression and PTSD were more prevalent amongst female adolescents compared with male adolescents, and the rate difference was similar to that observed in the general population. Most prevalence estimates remained stable over time, with the exception of ADHD and conduct disorder, of which the prevalence increased in the past decade. Results also highlighted the differential distribution of disease prevalence by age group (adolescents vs. adults) amongst people who experience incarceration. This finding is important because it suggests that a universal approach to resource planning and allocation might be inadequate for prison mental health care. Widening access to treatment services and improving the general quality of care for detained adolescents is crucial to prevent further criminality in adulthood, especially when considering that psychiatric disorders are modifiable risk factors for violent reoffending.

6.2.2 What is the predictive validity of a violence risk prediction tool (OxRec) across different geographical settings (Tajikistan and England)?

A prediction model for violent reoffending amongst people released from prison (OxRec) was validated in Tajikistan (Chapter 3) and England (Chapter 4). This particular model has the advantage of being transparent; both in the reporting of study methods and performance measures at the development and validation stages, and in terms of its published regression formula. It is also scalable, and thereby addresses many limitations of commonly used tools for violence risk. Following an incremental validation strategy—which took the form of a simple recalibration in England, and was supplemented by the selective re-estimation of coefficients for a single predictor (i.e. length of incarceration) in Tajikistan—both of these external validation cohort studies suggested that OxRec had good discrimination and calibration in predicting violent reoffending after release from prison. OxRec demonstrated robust predictive performance in these new geographical settings, despite considerable differences in social, cultural and economic contexts compared to the original development setting (Sweden). This is particularly noteworthy in Tajikistan given that no other validation study has been conducted in an LMIC to date, and that LMICs have been neglected in prison health research more broadly. The results suggest that OxRec, and other high-quality risk assessment tools developed in HICs, could be generalisable to other LMICs pending some operationalisation modifications (mostly to predictor definitions) to ensure that they are well adapted to the local context. Prior to the widespread adoption of OxRec in Tajikistan, England and other countries, feasibility and clinical impact studies should be undertaken to anticipate potential operational difficulties arising from its implementation. Such approach has been used for other risk assessment tools from the OxRisk initiative, spearheaded by the Forensic Psychiatry and Psychology group at the University of Oxford (Beaudry et al., 2022; Cornish et al., 2019; Forsman et al., 2022; Senior et al., 2020; Zhong, Yu, et al., 2021). Taken together, the findings of this research support the idea that OxRec has the potential to improve violence risk stratification in individuals who experience incarceration, and thereby optimise the allocation of effective interventions for violent reoffending.

6.2.3 What is the effectiveness of psychological interventions on recidivism after release from incarceration?

Chapter 5 evaluated the effectiveness of psychological interventions delivered in prisons and jails on recidivism following release. The generalisability of most published meta-analytic research on this topic is problematic due to the inclusion of studies of varying methodological quality and sample size. Previous reviews have also focused on specific types of interventions (e.g. CBT) or groups (e.g. individuals who use drugs). The purpose of this chapter was to provide robust and comprehensive effectiveness estimates. In order to do so, inclusion was limited to RCTs only. Overall, I found that the beneficial effects of psychological interventions in prison for recidivism had been previously overestimated, and that these were at best modest (OR = 0.72, 95% CI 0.56–0.92) based on the 29 eligible RCTs. I further tested the robustness of therapeutic effects by removing smaller trials (with less than 50 individuals in the intervention group), and these were no longer associated with decreased odds of reoffending (OR 0.87, 95% CI 0.68–1.11). Contrary to previous meta-analyses, CBT-based interventions were not effective in reducing recidivism (OR = 1.00, 95% 0.69–1.44). Rather, therapeutic communities had the strongest effects, although this was only based on evidence from two large included trials (OR = 0.64, 95% CI 0.46–0.91). Nevertheless, this result warrants further investigation to evaluate cost-effective interventions that extend continuity of care to community settings. These findings suggest that in general, widely implemented psychological interventions in prison require improvement in order to effectively reduce recidivism. This has important implications for future violence risk assessment strategies, as even the best prediction models will have very limited benefits if individuals identified as high risk are exposed to relatively ineffective therapies.

6.3 General strengths and limitations

The studies described in this thesis have various strengths and limitations, most of which have already been discussed comprehensively in the previous empirical chapters (Chapters 2–5). Thus, in the section that follows, I will focus on the overarching strengths and limitations with the aim to provide an unbiased and in-depth overview of the research that constitutes this thesis.

The methods employed have many strengths. I applied rigorous statistical methods—meta-analysis (meta-regression where possible) and validation approaches for predictive modelling—to address key evidence gaps in the fields of prison public health and violence risk prediction. The underlying methodology was developed according to best practice in systematic reviews (Higgins & Green, 2019) and clinical prediction modelling research (McLernon et al., 2022; Royston & Altman, 2013; Steyerberg, 2009; Steyerberg & Vergouwe, 2014; Van Calster et al., 2019). It has been described in detail throughout the various chapters to ensure the reliability and transparency of the research (Bell, 2017). I preregistered the search protocol for the systematic reviews (Pieper & Rombey, 2022), and developed extensive analysis plans for both validation studies. These were also validated by an independent ethical approval process, when appropriate. Preregistration ensures that the performed analyses are confirmatory (rather than exploratory) and outcome-independent (rather than outcome-dependent), which in turn reduces the risk of bias (e.g. selective reporting) (Hardwicke & Wagenmakers, 2022; Nosek et al., 2018).

A further advantage of preregistration is that it allows other researchers to evaluate the risk of bias independently, and adopt a critical perspective when evaluating the data and the resulting conclusions (Hardwicke & Wagenmakers, 2022).

Nevertheless, there are inherent limitations to this thesis, one of which being that I did not employ machine learning methods. The number of machine learning prediction model studies is increasing rapidly (Collins & Moons, 2019); yet, these are few and far between in prison public health and forensic psychiatry. As such, a systematic comparison of the predictive accuracy of regression-based prediction models and those based on machine learning is lacking in these fields of clinical medicine. At the start of my DPhil in October 2018 (Michaelmas 2018), I had aimed to compare and contrast, with a given data set, the ability of a Cox proportional-hazards model and a machine learning technique to predict violent reoffending amongst adolescents exposed to the criminal justice system. My original research plan also proposed to expand and enrich the discussion surrounding the use of machine learning to predict recidivism, taking into consideration of the risk factors that would have been identified in the prognostic model. The findings were expected to help identify young people at high risk of violent reoffending, and hence offer them better tailored interventions, and improve public health programmes to reduce youth crime.

These analyses were to be conducted in a large provincial data set of people with convictions in British Columbia (BC; Canada). The British Columbia Inter-Ministry Research Initiative (IMRI) combines records from publicly funded programmes of three independent provincial government entities: justice, health and social development (Rezansoff et al., 2013). Information sharing agreements between the BC ministries and Simon Fraser University (SFU) oversee the IMRI (Somers et al., 2016). Given that access to the database requires security clearance and is limited to on-site authorised users, it was estimated that this would require me to spend approximately one month performing data analyses under the supervision of Professor Julian M. Sommers at SFU. In January 2020, I visited our collaborators at SFU to conduct preliminary analyses and discuss my upcoming research fieldwork, which was originally planned for March 2020. However, my fieldwork was first postponed, and eventually cancelled due to the ongoing pandemic.

By then, I had already completed the reviews (Chapter 2 & 5), which did not necessitate ethical approval nor external data sources. Whilst developing alternative DPhil research projects with my supervisors and securing access to new data sources, I led on another review that aimed to synthesise the evidence on outbreaks of highly contagious diseases in prison to inform public health responses to COVID-19 (Beaudry et al., 2020). I was fortunate to gain access to two additional data sets from Tajikistan and England during the second wave of the pandemic, although these were comprised of adult subjects rather than adolescents. This meant that I had to broaden the scope of my DPhil, by focusing on adults who had experienced incarceration, rather than solely adolescents. Despite being unable to complete my initial research plans, and incorporate machine learning prediction algorithms in this body of work, it is anticipated that the present research will make significant contributions to enhancing the assessment and treatment of violence risk in prison populations across the globe.

Some principles of open science in mental health research were not followed in this thesis (Bell, 2017). I did not publish my statistical analysis code on a repository (e.g. GitHub) and some data (relating to validation samples from Tajikistan and England) could not be made publicly available for confidentiality reasons. This is to be expected when conducting research with criminal offence data, as the personal information of people with convictions is treated with the highest standards of confidentiality in most jurisdictions. Most empirical chapters have now been published in scientific journals; however, I did not consider publishing earlier versions of the articles as preprints (Bell, 2017), which could have accelerated the dissemination of my academic work, generated additional feedback and facilitated collaborations, especially in the context of the pandemic (Sarabipour et al., 2019).

Another consideration is that, although best practice was followed for the validation studies of OxRec, one of the tool developers (Prof Seena Fazel) was involved in the study design and the interpretation of results. This is noteworthy especially because there is evidence of authorship bias in existing validation studies of violence risk assessment tools (Singh et al., 2013). Authorship bias occurs when researchers who initially developed the model also validate it (rather than independent investigators) and it can result in performance overestimation (Boutron et al., 2010; Ochodo et al., 2013). For instance, a recent meta-analysis analysed separately independent validation studies and those in which tool developers were co-authors, with findings suggesting an overestimation of predictive performance in the latter (Fazel, Burghart, et al., 2022). This potential authorship effect was mitigated in that I conducted the analyses and first interpreted the results under the direct supervision of my other supervisor (Dr Rongqin Yu).

6.4 Implications for research, remaining evidence gaps and future directions

6.4.1 Psychiatric epidemiology

The literature on the prevalence of psychiatric disorders in adult prison populations has been extensively characterised (see for instance, Baranyi et al., 2019; Fazel, Hayes, et al., 2016; Fazel and Seewald, 2012; Fazel, Yoon, et al., 2017). Similar synthesis efforts were previously undertaken for adolescent criminal justice populations (Fazel, Doll, et al., 2008), but these prevalence estimates are now outdated due to a large amount of research being published in the last decade. Chapter 2 updates the evidence base on the prevalence of common and treatable mental disorders amongst adolescents in juvenile detention and correctional facilities. The results confirm the high burden of mental illness in adolescents who experience detention compared with their general population counterparts. Prevalence trends have direct implications for future mental health care provision, and most of these have remained stable over time. It can therefore be assumed that there is no pressing need for undertaking further prevalence studies (at least in HICs), but rather that future research efforts should be directed at developing evidence-based interventions to improve mental health outcomes.

However, additional prevalence research is needed in LMICs, as no primary studies from these countries were identified in the context of this thesis (Chapter 2). Various strategies could be employed in an attempt to obtain comprehensive prevalence estimates from LMICs. For instance, focus groups and interviews could be conducted with clinicians and justice professionals based in LMICs to adapt the current search strategy to better reflect the local terminology (e.g. borstal) (Atilola, 2021). This could potentially yield relevant new studies or those otherwise missed. Whilst global estimates are needed to reliably examine trends and anticipate future burden on health services (Frank et al., 2019), regional estimates are the most helpful for public health policymakers to establish mental health priorities (Neufeld, 2022; Signorini et al., 2017). Thus, future meta-analyses would also benefit from aggregating prevalence data by region or continent, as these findings should better reflect regional level needs, and in turn improve public policymaking based on local applicability to reduce mental health-related inequalities between countries (Lavis et al., 2004). This was not possible in the current review, given that there were not enough studies to aggregate them to the region/continent level.

The current prevalence of some psychiatric disorders (i.e. ADHD and conduct disorder) in the adolescent criminal justice population has increased over recent decades. Potential reasons for this have been discussed in Chapter 2, but these remain largely unclear and will require further study. Importantly, changes to ADHD prevalence estimates over time have not been mirrored in the general adolescent population (Thomas et al., 2015). Hence, it could be conceivably hypothesised that factors which have been known to increase the risk of involvement in the criminal justice system, such as difficulties in childhood/adolescent health and development, and others related to experiencing incarceration, may provide insight into this group-specific increase (Hughes et al., 2020). Increased awareness and screening of ADHD in correctional facilities could also have played a role. As for conduct disorder, the extent to which the prevalence has changed over time in the general adolescent population is still a matter of debate (Collishaw et al., 2004; Erskine et al., 2013; Fombonne, 1998), and this makes comparison to the trends observed in the adolescent criminal justice population more difficult (Fairchild et al., 2019). This might be further complicated by the fact that the prevalence could be overestimated in incarcerated young people owing to overdiagnosis of conduct disorder. Overdiagnosis due to overlapping symptoms (and behaviours) is a possibility, considering the significant comorbidity between conduct disorder and criminality (Copeland et al., 2007; Merten et al., 2017).

It is still not known if imprisonment throughout adolescence might aggravate preexisting mental health conditions or possibly cause new ones. There is some qualitative evidence for this in adult women (Harner & Riley, 2013), although findings from this thematic analysis should be interpreted with caution given that responses varied considerably between individuals. The association between incarceration and mood disorders, such as major depressive disorder and bipolar disorder, has previously been established in adult individuals using US data (i.e. the National Comorbidity Survey Replication) (Schnittker et al., 2012). Yet, these findings are yet to be replicated in other jurisdictions, despite this investigation being published nearly a decade ago. Further cohort studies using a within-individual design will need to be undertaken to investigate the potential association between incarceration and mental health outcomes amongst adolescents, and the possible underlying mechanisms.

6.4.2 Psychological interventions

Numerous large-scale systematic reviews have suggested that psychological interventions commonly delivered to people in prison, most notably CBT, are effective for reducing recidivism (see for instance Papalia et al., 2019). Chapter 5 contradicts partially the previous evidence, as it finds that the effects of treatments offered in prison for repeat offending have been overestimated in earlier meta-analytical studies. This represents a major implication for prison health research, similar to that observed in the literature on the efficacy of psychological interventions for adult depression (Cuijpers et al., 2010). Together, these results point towards a clear need for more effective interventions to reduce violent reoffending, which also requires considerable improvement to the general quality of efficacy trials in prisons and other closed settings.

Therapeutic communities and interventions that ensure continuity of care in community settings were identified as the most promising intervention types, albeit this finding being based on a very small number of trials. These interventions have two clear advantages over other types of treatment; they tend to be devised for delivery in custody settings and to target specific psychosocial needs for released individuals, such as substance misuse. Failure to meet basic needs on release including adequate medical treatment (both relating to physical and mental health), safe housing and employment opportunities contributes to repeat offending, with individuals eventually returning to their initial criminal behaviour as a way of supporting themselves and their families. This also increases the risk of all-cause morbidity and mortality for formerly imprisoned persons, notably with regards to drug overdose (Binswanger et al., 2013). Thus, transitional care interventions that directly address those needs, which have shown promising results in several vulnerable populations (for instance amongst formerly incarcerated adults with substance use disorders (Burns et al., 2022), and people with severe mental illness and a history of homelessness (Aubry et al., 2015)) should be prioritised for further investigation. Better integration of prison health care with wider public health structuring, and improved linkage of electronic health care records between prisons and community services is likely to facilitate such liaison efforts (Bellass et al., 2021).

The dearth of methodologically robust RCTs evaluating psychological therapies in custodial settings may be attributable to methodological obstacles (Quina et al., 2007). This is also likely linked to longstanding insufficient government investment in prison health care, particularly in LMICs, which might translate into difficulties to conducting clinical trials that adhere to the highest research standards and best practice ethical guidelines. Carrying out research in prisons and other closed settings may be further complicated by the fact that these are not inherently therapeutic environments, with various ethical, structural and security concerns at stake. Nevertheless, these potential challenges should not undermine research efforts, and thus there is an urgent need for additional work to improve existing procedural guidance on the ethical conduct of prison health research (Ako et al., 2020). Although some progress has been made in recent decades to close the ethical treatment gap (Byrne, 2005; Fox et al., 2011; Innes & Everett, 2008; Quina et al., 2007; Wakai et al., 2009), it has mostly been restricted to HICs, thereby further contributing to disparities in health research outputs between these countries and LMICs (Ako et al., 2020).

6.4.3 Risk assessment

Risk assessment tools are frequently employed in the criminal justice system. Such tools are intended to assist criminal justice professionals in making decisions at several points in the legal pathway about sentencing, supervision and treatment. When developed, validated and implemented according to best practice, prediction models should improve decision-making and subsequent criminal outcomes (Fazel, Sariaslan, et al., 2022). Yet in practice, many of these tools have important practical and methodological limitations (see Chapters 1 & 3–4, or alternatively Fazel, Burghart, et al., 2022 & Fazel, Sariaslan, et al., 2022 for more detail), and a small minority have been validated in settings that differ from the one in which they were initially derived. This is especially the case for LMICs, where to my knowledge, no risk prediction models for recidivism have been derived nor validated. Chapters 3 and 4 extend the current evidence by establishing the predictive accuracy of OxRec in two additional countries (Tajikistan and England), and their findings expand upon a previously reported external validation study in the Netherlands (Fazel et al., 2019). Overall, the reported accuracy, discrimination and calibration in these external samples provide strong evidence for the transportability of the model to samples other than the Swedish cohort used for the development of OxRec. Importantly, no systematic overestimation of risk was observed in any of these validations—a concern which would have important legal ramifications (Fazel, Sariaslan, et al., 2022).

Very few prediction models have been validated more than once in a population-based setting (Fazel, Burghart, et al., 2022), despite cardiovascular research indicating that single validations provide limited information on predictive performance (Wessler et al., 2021). To date, OxRec has been externally validated three times, and subsequent validations are underway in Finland, Norway and Estonia. Future research should strive to align with recent calls for more rigorous stewardship surrounding the construction and the validation of predictive algorithms (Adibi et al., 2020; Eaneff et al., 2020; Wessler et al., 2021). As previously suggested by Eaneff et al. (2020), this could take the form of an online repository for prediction models currently in use, whereby results, predictive measures and the final models would be reported in a transparent manner. Such resource would require regular maintenance, recalibration and updating (Wessler et al., 2021), but it is expected that the potential benefits would be worth the ensuing economic costs, time investment and resource implications (Guinney et al., 2017).

In a broader data sharing perspective, development and validation databases could also be openly shared in a secure and responsible manner to widen access to clinical data on people experiencing incarceration (Adibi et al., 2020). This would likely improve scientific rigour in the field by increasing accountability between researchers, and also enable them to gain a more detailed understanding of health and justice data through multidisciplinary and international partnerships (Guinney et al., 2017). Clinicians and policymakers would also benefit from more transparent approaches to prediction modelling, as these would facilitate more informed decision-making regarding the use of risk assessment tools, particularly with respect to evidence-based treatment allocation (Fazel, Sariaslan, et al., 2022).

Whether OxRec could enhance current risk assessment and management practices, and thereby contribute to reducing violent reoffending is an empirical question that remains to be addressed. Parallel to ongoing validation efforts, implementation studies will be needed to determine the feasibility, impact and utility of this tool in various criminal justice settings. Ethical (e.g. potential harms), economic (e.g. direct and indirect costs) and practical considerations (e.g. linkage to available treatment and interventions after being labelled as 'high risk' by the tool) are central to the utility testing and successful implementation of such models (Fazel, Sariaslan, et al., 2022). Those would be best assessed in RCTs, although such research designs are rare with regards to clinical prediction models. Cloud infrastructures could be a cost-effective alternative to standard randomisation procedures (Adibi et al., 2020), particularly in low-income settings where investment in prison health research is often suboptimal (Ako et al., 2020). For instance, eligible individuals who experience incarceration, justice professionals or clinicians, and custodial settings could be identified via the aforementioned sharing strategies and subsequently randomised to two or several approaches for them to be readily evaluated. This would allow for all processes and outcomes to be compared in real time, and thus facilitate rapid clinical translation if and when results are conclusive (Adibi et al., 2020).

The extent to which more complex methods for developing prediction models in criminal justice, such as machine learning, will improve their prognostic value is also an area for future research. In fact, it has been suggested that machine learning techniques (e.g. neural networks) could have an advantage over traditional statistical models when analysing very large datasets. As such, these methods could have the potential to uncover associations between predictor variables and a given outcome, which remain unaccounted for in traditional linear models, particularly where the number of predictors and their interactions are very large (Jordan & Mitchell, 2015; LeCun et al., 2015). Few investigations have undertaken such approach in criminal justice research (see for instance, Takahashi and Evans, 2018), and the conclusions arising from their analyses are highly limited by small sample sizes, which undermine the generasibility of their results (Banerjee et al., 2021; Navarro et al., 2021). Moreover, evidence from other branches of science, such as oncology (Dhiman et al., 2022), has not been conclusive, with no improved benefits of machine learning over traditional statistical models (in terms of performance) (Chen & Asch, 2017; Christodoulou et al., 2019; Shillan et al., 2019; Shung et al., 2019; Song et al., 2021; W. Wang et al., 2020). Nevertheless, additional research on machine learning methods in criminal risk prediction is warranted, as improvements to performance could result in added benefits for individuals who are involved in the justice system. These potential benefits would need to be weighed against the additional statistical complexity introduced by these novel approaches, which may result in longer administration times and further complicate risk communication. Overfitting, which occurs when too many predictors or features are added relative to the sample size, and general lack of transparency in artificial intelligence prediction models would also require consideration. These methodological concerns impede on the interpretability of identified associations between the outcome and predictors, and this could ultimately prevent against successful translation into clinical practice (Collins & Moons, 2019).

An alternative approach that has been pursued in cardiovascular prediction research to improve predictive performance is to derive sex- and age-specific models, which are then calibrated to specific risk regions. Such approach (SCORE2; SCORE2 working group and ESC Cardiovascular risk collaboration, 2021) has been found to improve risk discrimination compared to a previous, simpler model (SCORE; Conroy et al., 2003) with an increased ability to identify individuals at higher risk of developing cardiovascular diseases (CVD), particularly amongst the younger population. The rationale for this model revision was that its first iteration lead to risk misestimation in certain circumstances due to varying CVD risk levels between subgroups and regions (SCORE2 working group and ESC Cardiovascular risk collaboration, 2021). Considering that previous research clearly suggests that there are age-, sex- and region-based differences in the incidence of violent reoffending and related risk factor distributions (similar to that observed with CVD) (Chang et al., 2015; Fazel, Chang, et al., 2016; Sariaslan et al., 2020), future work should be undertaken to investigate whether a similar methodology could be used to improve and refine OxRec. Further, including age-specific interactions for some predictor variables of which the association with violent reoffending varies with age (e.g. psychiatric disorders) should also be considered (Chang et al., 2015). This also relates to Chapter 2 which further demonstrated that the prevalence of specific mental disorders differs between age groups (adolescents vs. adults). If OxRec was refined using registry data based on this procedure for derivation and that new epidemiological data with age- and sex-specific information were made available, the model could be easily updated at a country-level (both in HICs and LMICs based on Chapters 3 & 4) to reflect future violent reoffending incidence and risk factor profiles (SCORE2 working group and ESC Cardiovascular risk collaboration, 2021).

As for treatment allocation strategies that are informed by prediction models, further empirical work is required to determine to what degree people released from prison actually benefit from these individualised treatment decisions. Prediction research related to COVID-19 has suggested that individuals who are most likely to benefit from effective interventions should be prioritised for treatment, rather than those who are simply at higher risk of the outcome (violent reoffending, in this case) (de Jong et al., 2022). To reliably identify people who stand to benefit the most from these treatments, however, sophisticated models that forecast (counterfactual) individual outcomes for all available interventions options and their potential consequences are required (Hoogland et al., 2021; Nguyen et al., 2020). This would also involve estimating differential treatment effects for every given individual (with vs. without treatment) (Hoogland et al., 2021), and therefore whether this is feasible in prisons and jails is unclear at present.

6.5 Implications for policy and practice

Violent reoffending is a pressing public health issue that not only impacts individuals—whether they be victims or perpetrators—but also communities and society. This thesis raises a number of policy and practice implications for better addressing violent reoffending amongst people released from prison in a public health perspective. Thus, the following section details key considerations that stem from the findings of each empirical chapter.

Chapter 2 demonstrates that the adolescent criminal justice population is characterised by elevated psychiatric morbidity, as indicated by higher prevalence rates for multiple mental disorders compared with those observed in the general adolescent population. A wealth of research suggests that poor mental health and exposure to the juvenile justice system increase the likelihood of future offending, yet access to adequate mental health care in correctional facilities is rare in LMICs and highly variable between HICs (see for instance Hughes et al., 2020). This should be noted by policymakers in order to rethink legal pathways for children and adolescents who commit crimes, particularly those who have mental health conditions and those who fall under the age of criminal responsibility. Future policies should prioritise alternatives to prosecution in the criminal justice system, such as diversion programmes and restorative actions, and prioritise immediate linkage to high-quality community psychiatric care or inpatient services, when needed (Lucas & Staines, 2022).

Chapter 5 finds that widely implemented psychological interventions in prison have little to no beneficial effects in reducing recidivism. Notably, few robust trials were identified in this systematic review, and none were from LMICs. Therefore, there is a clear need for investment in research to develop new effective interventions and to improve existing ones that have shown promising preliminary results, particularly therapeutic communities. Sex-specific interventions should also be further evaluated, as these were found to be effective in subgroup analyses. Such research is especially needed in LMICs, but this will also require considerable advocacy work to mitigate societal stigma surrounding prison populations in those settings, as this has historically translated into a general reluctance from governments to invest in prison health research (Ako et al., 2020). This area of research could inform much-needed policy and practice changes in the provision of psychological treatment within prisons, and improve transition and community-based programmes for those leaving prison (Scheyett, 2022).

Chapters 3 and 4 tested the external validity of OxRec and found good predictive performance in Tajikistan and England. This underlines the possibility of integrating high-quality risk assessment tools in current clinical decision-making practices to guide the allocation of limited resources within prisons, such as effective psychological interventions, and reduce costs. Prediction models could also be used to inform decarceration efforts and redirect individuals at lower risk of violent reoffending to community-based sentencing programmes (Fazel, Sariaslan, et al., 2022). In recent years, strategies that reduce incarceration of people who inject drugs and cost-effective interventions to diagnose, treat and prevent blood-borne viruses have yielded some benefits in several LMICs from Eastern Europe and Central Asia, particularly in Azerbaijan (Altice et al., 2016; WHO, 2019). Similar approaches to other serious and often co-occurring health problems that disproportionately affect individuals interfacing with the criminal justice system should be considered (Fazel, Bromberg, et al., 2022), as the provision of health care is the most effective method for reducing recidivism and enhancing prison public health, especially when embedded within a harm reduction perspective (Altice et al., 2016). Potential savings from effective decarceration strategies could then be reinvested by governments into prison health research funding and towards the costs of employing additional health professionals in correctional facilities to improve service provision for mental health. To facilitate these reforms, a task force of clinicians, researchers, ethicists, legal experts, individuals with lived experience of incarceration and other stakeholders could be convened to advise on new developments in risk prediction. How best to translate clinical evidence into integrated policy and system approaches to violence prevention, identification and rehabilitation also need to be carefully considered (Sanci, 2019; Wathen & MacMillan, 2018).

6.6 Conclusions

Despite being largely predictable and preventable, violent reoffending amongst people released from prison remains a major public health issue worldwide. Novel approaches to violence risk assessment have been developed and trialed in this key population over the past decades. Yet, the field of prediction modelling and its applications in prison health research are marked by important methodological limitations, and LMICs have historically been overlooked. Moreover, the beneficial consequences of being classified as high risk for violent reoffending are few and far between, with meta-analytical research suggesting that widely implemented psychological interventions in prison have at best modest effects on recidivism. This translates into a reluctance to implement these tools in criminal justice settings, a concern with which other prediction models in cardiovascular and cancer domains are not necessarily met. Further, access to mental health care in prisons remains scarce, notwithstanding high prevalence of mental illnesses and substance use disorders in adolescent and adult prison populations, and their importance as modifiable risk factors for violent reoffending. Adequately developed and validated, high-quality violence prediction models (such as OxRec) have the potential to enhance risk stratification, and in turn inform decarceration efforts and linkage to effective treatment. However, large clinical trials in prisons are needed to improve existing psychological interventions and develop new ones. Such approach is likely to generate considerable savings that could be reinvested within the criminal justice system to improve the general quality of psychiatric care in correctional facilities.

Appendices

A

Appendix A: Prevalence of mental disorders in detained adolescents

A.1 Quality assessment

(1) Representativeness of the sample

1 point: Stratified random, random, population or systematic sampling.

0 point: Convenience sampling.

(2) Sample size:

1 point: Sample size equal to or greater than 100 participants.

0 point: Sample size less than 100 participants.

(3) Participation:

a. Reporting of participation rate or non-response analysis

1 point: Yes.

0 point: No.

b. Satisfactory rate of participation (if refusal rate not reported, non-response rate used)

1 point: Participation rate was equal or higher than 80%.

0 point: Participation rate was lower than 80%, or not reported.

(4) Validity of mental disorder diagnosis:

1 point: Psychiatrist(s)/psychologist(s) made the mental disorder diagnoses.

0 point: Trained interviewer(s) made the mental disorder diagnoses.

(5) Quality of descriptive statistics:

1 point: Reporting of the descriptive statistics to describe the sample included age AND at least one other socio-demographic or criminal characteristic.

0 point: Descriptive statistics were not reported or were incomplete.

Scoring:

1–2 Low

3–4 Medium

5–6 High

Note. This scale is based on a modified version of the Newcastle-Ottawa Scale found in Baranyi et al., 2018.

Table A.1: Quality scoring of included samples, 1966–2019

Study	Representativeness	Sample size	Reporting of participation rate	Satisfactory rate of participation	Validity of diagnosis	Quality of statistics	Total score	Quality score
Abram et al., 2004 (M)	1	1	1	1	0	1	5	High
Abram et al., 2004 (F)	1	1	1	1	0	1	5	High
Abrantes et al., 2005 (M)	1	1	0	0	0	1	3	Medium
Abrantes et al., 2005 (F)	1	0	0	0	0	1	2	Low
Aebi et al., 2015	1	1	1	1	1	1	6	High
Aebi et al., 2016	1	1	1	1	1	1	6	High
Aida et al., 2014	0	1	1	1	1	1	5	High
Atkins et al., 1999 (M)	1	0	1	1	0	1	4	Medium
Atkins et al., 1999 (F)	1	0	1	1	0	1	4	Medium
Bolton, 1976 (M)	1	1	0	0	0	1	3	Medium
Bolton, 1976 (F)	1	1	0	0	0	1	3	Medium
Chiles et al., 1980 (M)	1	0	1	0	0	0	2	Low
Chiles et al., 1980 (F)	1	0	1	0	0	0	1	Low
Chitsabesan et al., 2006 (M)	1	1	1	1	1	1	6	High
Chitsabesan et al., 2006 (F)	1	0	1	1	1	1	5	High
Colins et al., 2009	1	1	1	1	0	1	5	High
Dimond and Misch, 2002	1	0	1	1	1	0	4	Medium
Dixon et al., 2004	1	1	1	1	1	1	6	High
Dória et al., 2015	1	0	0	0	0	1	2	Low
Duclos et al., 1998 (M)	1	0	1	0	0	1	3	Medium
Duclos et al., 1998 (F)	1	0	1	0	0	1	3	Medium
Ghanizadeh et al., 2012	0	1	1	1	0	1	4	Medium
Gonzalvo, 2002	1	0	1	1	1	1	5	High
Gosden et al., 2003	1	1	1	0	1	1	5	High
Gretton and Clift, 2011 (M)	1	1	1	1	0	1	5	High
Gretton and Clift, 2011 (F)	1	0	1	1	0	1	4	Medium

Table A.1: Quality scoring of included samples, 1966–2019 (continued)

Study	Representativeness	Sample size	Reporting of participation rate	Satisfactory rate of participation	Validity of diagnosis	Quality of statistics	Total score	Quality score
Guebert and Olver, 2014 (M)	0	1	0	0	1	1	3	Medium
Guebert and Olver, 2014 (F)	0	0	0	0	1	1	2	Low
Hamerlynck et al., 2007	1	1	1	1	0	1	5	High
Harzke et al., 2012 (M)	1	1	1	1	0	0	4	Medium
Harzke et al., 2012 (F)	1	1	1	1	0	0	4	Medium
Hollander and Turner, 1985	1	1	1	1	1	1	6	High
Indig et al., 2009 (M)	1	1	1	1	1	1	6	High
Indig et al., 2009 (F)	1	1	0	1	1	1	5	High
Karnik et al., 2010 (M)	1	1	1	1	0	1	5	High
Karnik et al., 2010 (F)	1	1	1	1	0	1	5	High
Kashani et al., 1980 (M)	1	0	0	0	1	1	3	Medium
Kashani et al., 1980 (F)	1	0	0	0	1	1	3	Medium
Kim et al., 2017	1	1	1	1	1	1	6	High
Köhler et al., 2009	1	0	1	1	1	0	4	Medium
Kuo et al., 2005 (M)	1	0	1	0	0	0	2	Low
Kuo et al., 2005 (F)	1	0	1	0	0	0	2	Low
Lader et al., 2000 (M)	1	1	1	1	1	0	5	High
Lader et al., 2000 (F)	1	1	1	1	1	0	5	High
Lederman et al., 2004	1	1	1	0	0	1	4	Medium
Lennox et al., 2013	1	1	1	1	0	1	5	High
Lindblad et al., 2015	0	1	1	1	1	1	5	High
Mitchell and Shaw, 2011	1	1	1	1	0	1	5	High
Nicol et al., 2000	1	0	1	0	0	0	2	Low
Pliszka et al., 2000 (M)	1	0	1	1	0	1	4	Medium

Table A.1: Quality scoring of included samples, 1966–2019 (continued)

Study	Representativeness	Sample size	Reporting of participation rate	Satisfactory rate of participation	Validity of diagnosis	Quality of statistics	Total score	Quality score
Pliszka et al., 2000 (F)	1	0	1	1	0	1	4	Medium
Robertson and Husain, 2001 (M)	1	1	0	0	0	1	3	Medium
Robertson and Husain, 2001 (F)	1	0	0	0	0	1	2	Low
Ruchkin et al., 2002	1	1	1	1	1	1	6	High
Schorr et al., 2019	1	0	1	1	1	1	5	High
Shelton, 1998 (M)	1	1	1	1	0	1	5	High
Shelton, 1998 (F)	1	0	1	1	0	1	4	Medium
Sørland and Kjelsberg, 2009	1	0	1	1	0	1	4	Medium
Teplin et al., 2002 (M)	1	1	1	1	0	1	5	High
Teplin et al., 2002 (F)	1	1	1	1	0	1	5	High
Ulzen et al., 1998 (M)	0	0	1	1	0	1	3	Medium
Ulzen et al., 1998 (F)	0	0	1	1	0	1	3	Medium
Vreugdenhil et al., 2004	1	1	1	0	1	1	5	High
Waite D., 2002 (M)	1	1	1	1	0	1	5	High
Waite D., 2002 (F)	1	1	1	1	0	1	5	High
Wasserman et al., 2002	1	1	1	1	0	1	5	High
Yoshinaga et al., 2004 (M)	1	0	1	1	1	1	5	High
Yoshinaga et al., 2004 (F)	1	0	1	1	1	1	5	High
Zhou et al., 2012	1	1	1	1	1	1	6	High

Note. F = Females; M = Males.

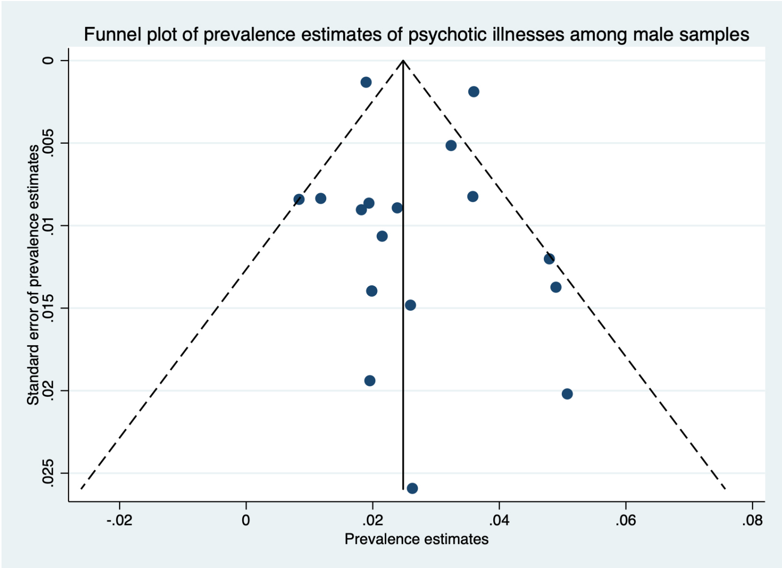


Figure A.1: Funnel plot of current psychotic illnesses prevalence estimates against standard errors (for male samples), 1966–2019. Egger's test suggested no significant bias in the male samples reporting prevalence of psychotic illnesses (Coef. = 0.322, SE = 0.662; t = .49, p= .63).

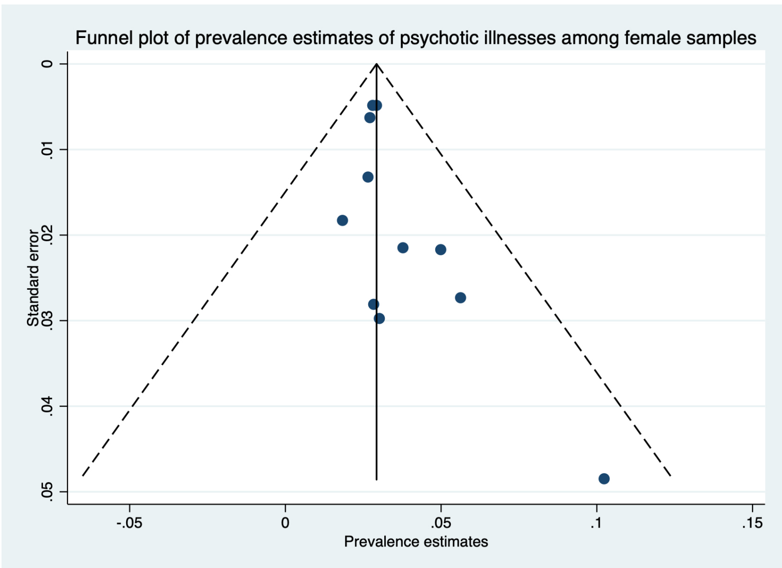


Figure A.2: Funnel plot of current psychotic illnesses prevalence estimates against standard errors (for female samples), 1966–2019. Egger's test suggested no significant bias in the female samples reporting prevalence of psychotic illnesses (Coef. = 0.605, SE = 0.289; t = 2.09 p = .07).

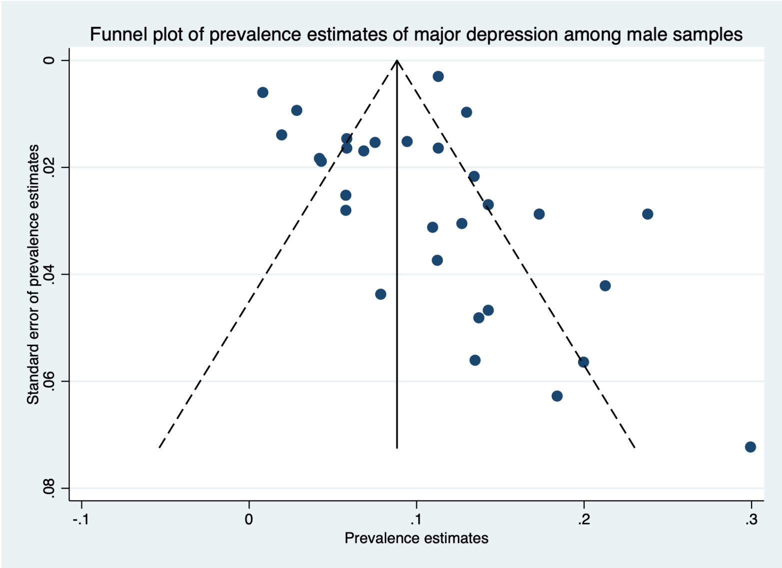


Figure A.3: Funnel plots of current major depression prevalence estimates against standard errors (for male samples), 1966–2019. Egger's test suggested no significant bias in the male samples reporting prevalence of major depression (Coef. = 0.261, SE = 0.986; t = .26, p = .79).

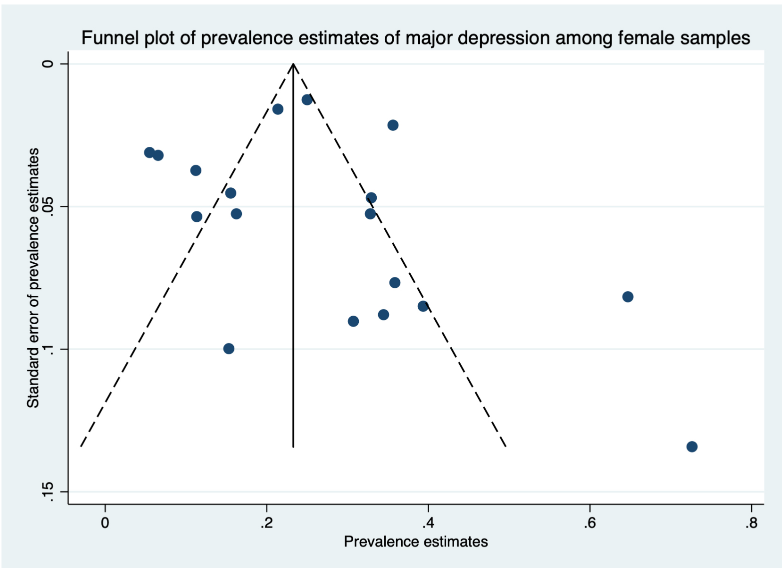


Figure A.4: Funnel plots of current major depression prevalence estimates against standard errors (for female samples), 1966–2019. Egger's test suggested no significant bias in the female samples reporting prevalence of major depression (Coef. = 0.664, SE = 1.283; t = .52, p = .61).

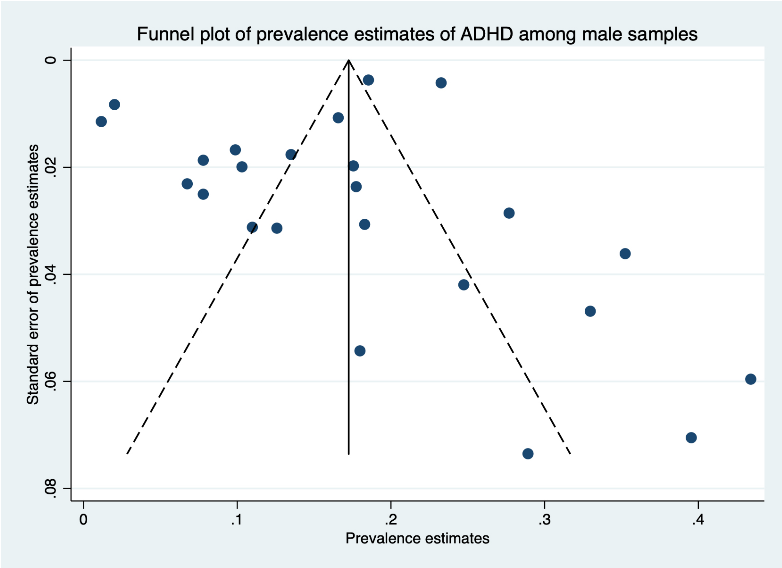


Figure A.5: Funnel plots of current attention deficit hyperactivity disorder (ADHD) prevalence estimates against standard errors (for male samples), 1966–2019. Egger's test suggested no significant bias in the male samples reporting prevalence of ADHD (Coef. = -1.481, SE = 1.824; t = -.81, p = .43).

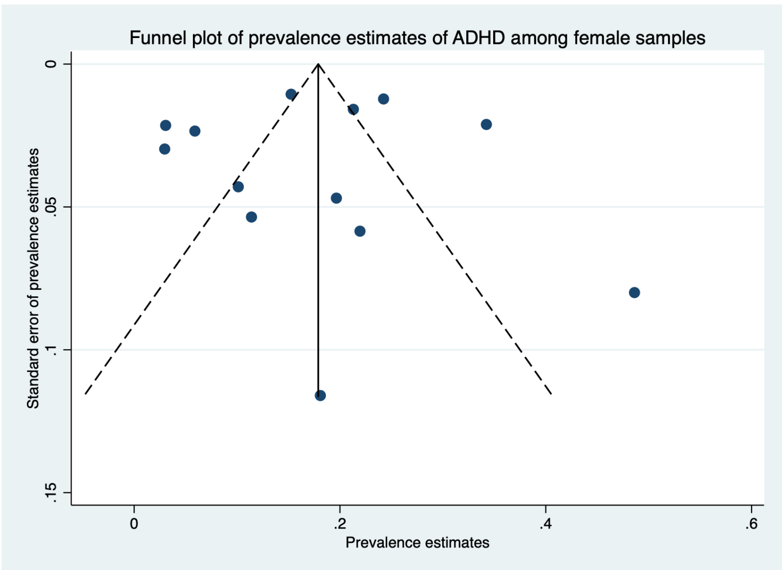


Figure A.6: Funnel plots of current attention deficit hyperactivity disorder (ADHD) prevalence estimates against standard errors (for female samples), 1966–2019. Egger's test suggested no significant bias in the female samples reporting prevalence of ADHD (Coef. = -0.563, SE = 2.211; t = -0.25, p = .80).

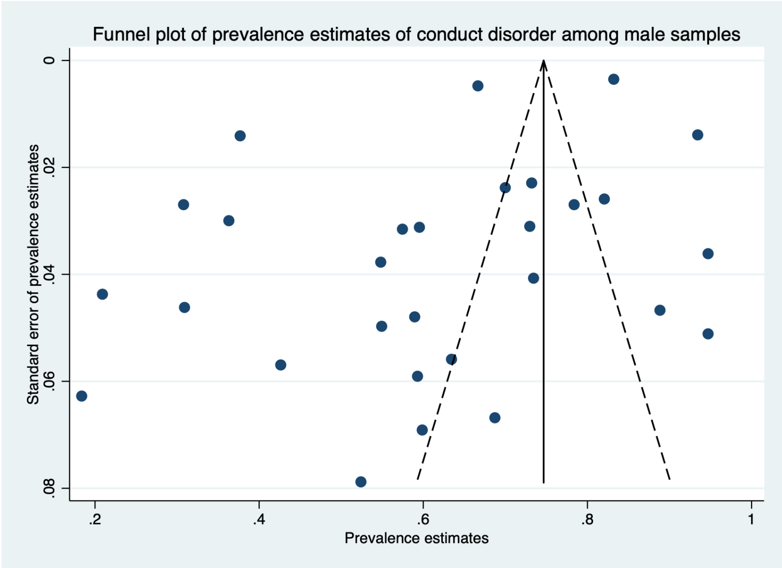


Figure A.7: Funnel plots of any lifetime conduct disorder prevalence estimates against standard errors (for male samples), 1966–2019. Egger's test suggested a significant bias in the male samples reporting prevalence of conduct disorder (Coef. = -4.984, SE = 2.128; t = -2.34, p = .03).

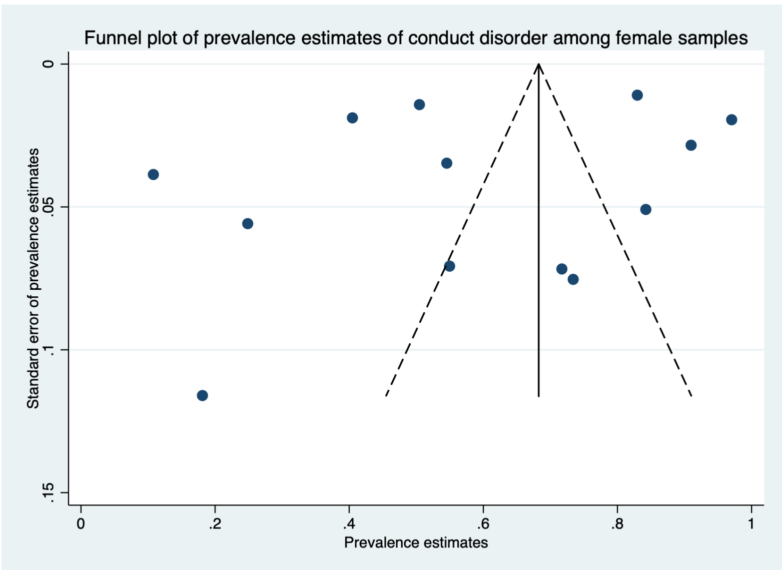


Figure A.8: Funnel plots of any lifetime conduct disorder prevalence estimates against standard errors (for female samples), 1966–2019. Egger's test suggested no significant bias in the female samples reporting prevalence of conduct disorder (Coef. = -4.378, SE = 4.703; t = -.930, p = .37).

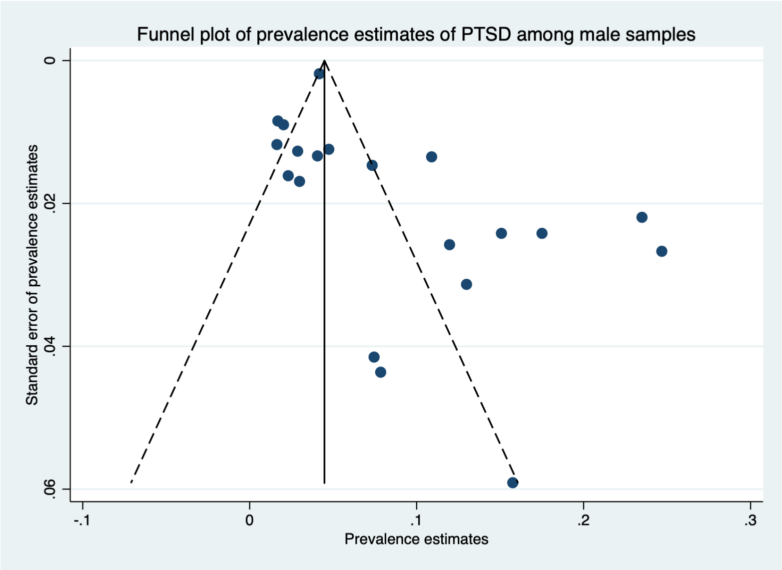


Figure A.9: Funnel plots of current posttraumatic stress disorder (PTSD) prevalence estimates against standard errors (for male samples), 1966–2019. Egger's test suggested a significant bias in the male samples reporting prevalence of PTSD (Coef. = 2.322, SE = 0.927; $t = 2.51$, $p = .02$).

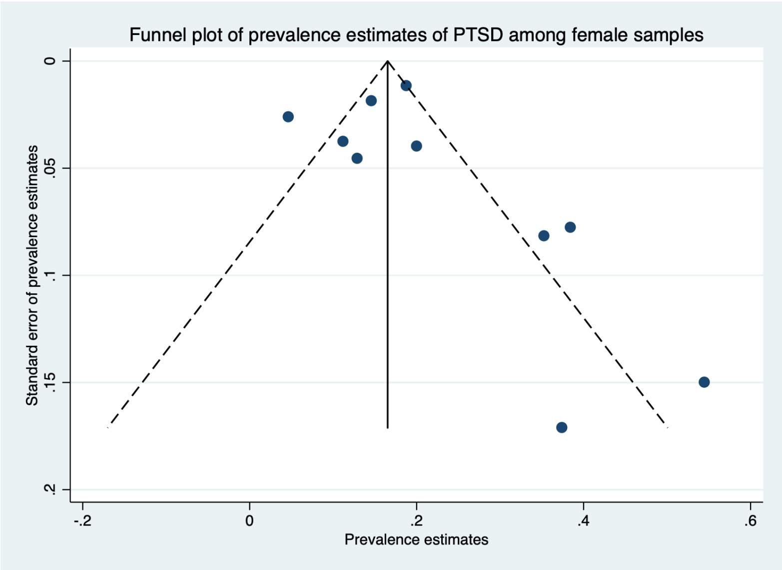


Figure A.10: Funnel plots of current posttraumatic stress disorder (PTSD) prevalence estimates against standard errors (for female samples), 1966–2019. Egger's test suggested no significant bias in the female samples reporting prevalence of PTSD (Coef. = 0.999, SE = 1.186; $t = .84$ $p = .42$).

B

Appendix B: OxRec external validation in Tajikistan

B.1 Study protocol

This project aims to examine the predictive validity of OxRec in Tajikistan. More specifically, we will test how well the OxRec can predict violent and any crime committed within one year (and if possible two years) of release from Tajikistan prisons. We will use routinely collected data including crime, mental disorders, and demographic factors such as gender, age, and employment (see Table 1.2). Researchers in Tajikistan will follow up a group of released prisoners for one year (and if possible two years). They will collect data on the following predictor variables: sex, age, length of incarceration, violent index offence, previous violent crime, civil status, educational level, employment before incarceration, disposable income, a diagnosis of alcohol use disorder or equivalent, a diagnosis of drug use disorder or equivalent, a diagnosis of any severe mental illness (schizophrenia-spectrum disorder or bipolar disorder), a diagnosis of any mental disorder (or equivalent using valid proxies for mental disorder). In addition, they will collect data on the outcome variables: Any criminal/violent reoffending within one year of follow up and if possible within two years. Colleagues at the Institute for International Health and Education (IIHE) are working in Tajikistan and they will facilitate the data collection process and one part time local coordinator and three local project assistants will be hired to collect the data from three prisons and jail.

After the data are collected, researchers at the University of Oxford will perform statistical analyses, following similar procedure as the Dutch OxRec validation study (Fazel et al., 2019), and write up research reports. The public benefits may include improved violence risk assessment in prisoners and targeting of appropriate interventions, by identifying individuals who are at the highest risk of re-offending and most in need of interventions, particularly for drug and alcohol use disorders, to reduce future criminality. More specifically, if the tool is validated, the risk calculator can be put online and used in Tajikistan by criminal justice professionals and healthcare staff. For instance, the tool could be used by prison health care to help guide treatment of prisoners before their release and community linkage, especially for those who need additional substance misuse treatment on release. It can also assist case workers in planning sentencing and release arrangement. In addition, as our tool does not require any training, any health-care or criminal justice professional such as judge, probation officers, psychiatrists, nurses could use them to assist in decisions on the timing of parole and conditions associated with it. Findings will be published in peer reviewed scientific journals, and presented at relevant academic conferences.

Table B.1: Variable definitions

Variable	**Sweden**	**Tajikistan**
Sex	Assigned at birth	Same definition
Age	Age at release from prison	Same definition
Immigrant status	First or second generation immigrants (self or either parent born outside of Sweden)	Not included
Length of incarceration	Duration of incarceration for most recent offence	Same definition

Table B.1: Variable definitions (continued)

Variable	Sweden	Tajikistan
Violent index offence	Most recent offence was homicide, assault, robbery, arson, any sexual offence (rape, sexual coercion, child molestation, indecent exposure, or sexual harassment), illegal threats, or intimidation	Same definition
Previous violent crime	Any conviction for a violent offence previous to most recent offence (i.e. before index offence)	Any incarceration for a violent offence as defined above, previous to most recent offence
Civil status	Unmarried vs. other (At imprisonment. Other includes married, cohabiting, divorced, and widowed)	Unmarried vs. other (At imprisonment. Other includes married, married more than once, and divorced)
Education	Lower secondary, upper secondary, post-secondary	None/high school (uncompleted), high school (completed), university (uncompleted), university (both uncompleted and completed)
Employment	Employed at incarceration (worked for at least 4 hours [based on their income information] during November before incarceration)	Employed at incarceration (including Tajik labour migrants)
Income	Negative/Zero/Low/ Medium/High (Low: $< 20^{\text{th}}$ percentile, Medium: 20-80$^{\text{th}}$ percentile, High: $> 80^{\text{th}}$percentile)	Low (below or equal to the international poverty line for low-income countries [1.90$ per person per day]) vs. stable (above this line)
Neighbourhood deprivation	Principal components analysis of: mean disposable income, % welfare recipients, % unemployed, % divorced individuals, % with only primary school qualifications, % of immigrants (defined as individuals who were not born in Sweden), residential mobility rate, and crime rate.	Not included
Alcohol misuse	Diagnosis of alcohol use disorder (lifetime: before or during incarceration—ICD-8: 291, 303; ICD-9: 291, 303, 305A; ICD-10: F10)	Alcohol misuse (self-reported)

Table B.1: Variable definitions (continued)

Variable	**Sweden**	**Tajikistan**
Drug misuse	Diagnosis of drug use disorder (lifetime: before or during incarceration—ICD-8: 304; ICD-9: 292, 304, 305 excl. 305A; ICD-10: F11-F19)	Drug misuse (self-reported)
Any mental disorder	Diagnosis of any mental disorder excluding substance use disorders (lifetime: before or during incarceration)	GHQ-12 total score $\geq$ 21
Any severe mental disorder	ICD diagnosis of schizophrenia-spectrum or bipolar disorder (lifetime: before or during incarceration)	SRQ total score $\geq$ 16

Note. GHQ = General Health Questionnaire; SRQ = Self-Reporting Questionnaire

Table B.2: Baseline characteristics of the Tajik sample compared with those of the Swedish sample (with adapted definitions)

Variable	Tajik sample (n = 970)	Swedish sample (n = 37,100)
Sex		
Male	846 (87%)	93%
Female	124 (13%)	7%
Age	Median 35	Median 36
	IQR 28–43	IQR 27–46
Length of incarceration		
$<$ 66 months	701 (73%)	$<$ 6 months — 69%
66–72 months	47 (5%)	6–12 months — 16%
72–84 months	64 (7%)	12–24 months — 10%
$\geq$ 84 months	158 (16%)	$\geq$ 24 months — 4%
Violent index offence	608 (63%)	38%
Previous imprisonment	105 (11%)	**Previous violent crime** — 53%
Civil status		
Other	578 (60%)	35%
Unmarried	392 (40%)	65%
Education		
None/high school (uncompleted)	57 (6%)	$<$ 9 years — 48%
High school (completed)	784 (81%)	9–11 years — 46%
University	129 (13%)	$\geq$ 12 years — 6%
Employment		
Unstable	481 (50%)	Unemployed — 75%
Stable	489 (50%)	Employed — 25%
Income		
Low	423 (44%)	Negative — $<$ 1% Zero — 6% Low — 53%
Stable	547 (56%)	Medium — 40% High — 1%
Alcohol use	358 (37%)	22%
Drug use	84 (9%)	23%
Any mental disorder	470 (48%)	22%
Any severe mental disorder	44 (5%)	3%

Note. Data are median (IQR) or n (%). Income was calculated using the international poverty line for low-income countries ($1.90 US per day).

Table B.3: Summary of outcomes in the Tajik sample and other OxRec samples

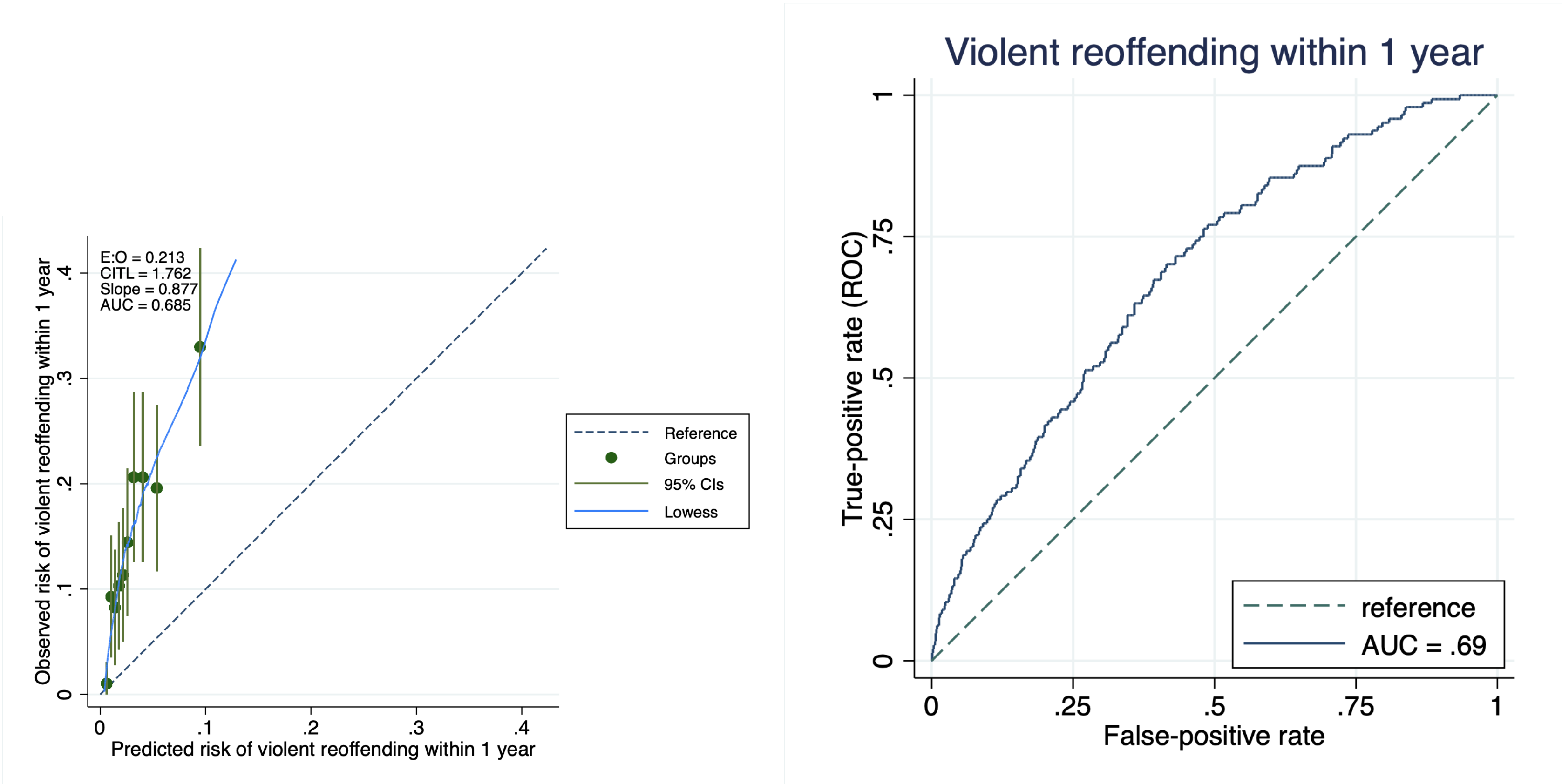
Outcomes	**Tajikistan** n = 970	**Netherlands** n = 9,072	**Sweden** n = 37,100
1 year violent reoffending	15%	8%	12%
1 year any reoffending	15%	28%	44%

Table B.4: Calibration performance measures for the uncalibrated OxRec model

	Observed number of events	**Expected number of events (uncalibrated)**	**Ratio, Expected: Observed (uncalibrated) (95% CI)**	**Ratio of crude event rates, Sweden: Tajikistan**	**Ratio, Expected: Observed (recalibrated) (95% CI)**	**Brier score**
1 year violent reoffending	144	31	0.21	1.25	1.09	0.12

Table B.5: Risk factors included in the final recalibrated model and their hazard ratios

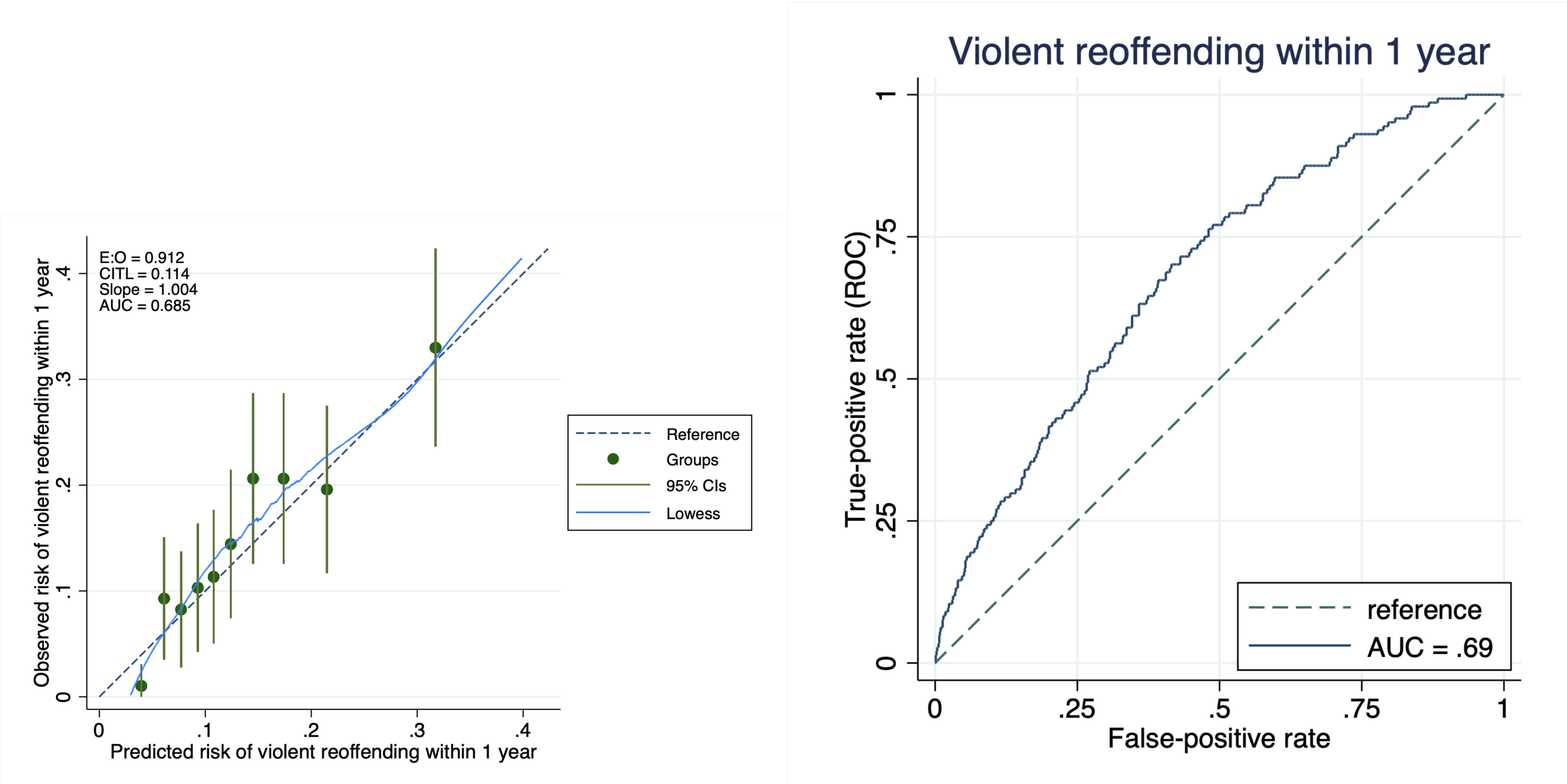
Variable	**Multivariable adjusted hazard ratio (95% CI)**	
	Tajikistan	**Sweden**
Sex (female)	0.29 (0.12–0.68)	0.51 (0.45–0.57)
Age	0.99 (0.97–1.01)	0.84 (0.83–0.85)
Immigrant		0.97 (0.92–1.02)
Length of incarceration		
$<$ 66 months	1	$<$ 6 months — 1
66–72 months	1.13 (0.51–2.48)	6-12 months — 0.85 (0.81–0.90)
72–84 months	1.82 (1.12–3.00)	12–24 months — 0.69 (0.63–0.75)
$\geq$ 84 months	0.50 (0.29–0.85)	$\geq$ 24 months — 0.55 (0.48–0.64)
Violent index offence	4.40 (2.58–7.50)	1.53 (1.46–1.59)
Previous imprisonment	1.89 (1.22–2.92)	Previous violent crime — 2.41 (2.29–2.54)
Civil status (unmarried)	0.85 (0.57–1.27)	1.08 (1.02–1.15)
Highest education		
None/high school (uncompleted)	1	$<$ 9 years — 1
High school (completed)	1.06 (0.54–2.07)	9–11 years — 0.83 (0.79–0.87)
University	0.81 (0.33–2.01)	$\geq$ 12 years — 0.65 (0.57–0.75)
Employment	0.77 (0.51–1.17)	0.68 (0.63–0.72)
Disposable income		
Low	1	Negative — 1 Zero — 1.69 (1.11–2.57) Low — 1.45 (0.96–2.19)
Stable	1.08 (0.71–1.64)	Medium — 1.45 (0.96–2.19) High — 1.57 (0.92–2.67)
Neighbourhood deprivation		1.03 (1.01–1.04)
Alcohol use disorder	1.22 (0.85–1.75)	1.41 (1.33–1.49)
Drug use disorder	2.46 (1.61–3.78)	1.51 (1.44–1.59)
Any mental disorder	1.35 (0.96–1.88)	1.09 (1.03-1.15)
Any severe mental disorder	2.56 (1.13–5.78)	1.10 (0.99–1.22)



(a) Calibration plot

(b) ROC curve

Figure B.1: Model performance after simple validation. AUC = area under the curve; CITL = calibration-in-the-large; E:O = expected to observed ratio; ROC = receiver operating characteristic.



(a) Calibration plot **(b)** ROC curve

Figure B.2: Model performance after recalibration. AUC = area under the curve; CITL = calibration-in-the-large; E:O = expected to observed ratio; ROC = receiver operating characteristic.

C

Appendix C: OxRec external validation in England

Table C.1: Variable definitions

Variable	**Sweden**	**England**
Sex	Assigned at birth	Gender (as entered in PINs)
Age	Age at release from prison	Age at release from prison, rounded down to full years
Immigrant status	First or second generation immigrants (self or either parent born outside of Sweden)	Nationality (as entered in PINs)
Length of incarceration	Duration of incarceration for most recent offence	Same definition
Violent index offence	Most recent offence was homicide, assault, robbery, arson, any sexual offence (rape, sexual coercion, child molestation, indecent exposure, or sexual harassment), illegal threats, or intimidation	Same definition
Previous violent crime	Any conviction for a violent offence previous to most recent offence (i.e. before index offence)	Same definition
Civil status	Unmarried vs other (At imprisonment. Other includes married, cohabiting, divorced, and widowed)	Same definition
Education	Lower secondary, upper secondary, post-secondary	Not included
Employment	Employed at incarceration (worked for at least 4 hours [based on their income information] during November before incarceration)	Last non-blank occupation before release (taken from NICHE Custody)
Income	Negative/Zero/Low/ Medium/High (Low: $< 20^{\text{th}}$ percentile, Medium: 20-80^{th} percentile, High: $> 80^{\text{th}}$percentile)	Not included
Neighbourhood deprivation	Principal components analysis of: mean disposable income, % welfare recipients, % unemployed, % divorced individuals, % with only primary school qualifications, % of immigrants (defined as individuals who were not born in Sweden), residential mobility rate, and crime rate.	Last valid residential or family home address before release date, passed through postcode-LSOA calculator and 2019 IMD index quintile

Table C.1: Variable definitions (continued)

Variable	**Sweden**	**England**
Alcohol misuse	Diagnosis of alcohol use disorder (lifetime: before or during incarceration—ICD-8: 291, 303; ICD-9: 291, 303, 305A; ICD-10: F10)	Remarks contain “alcoh” (taken from NICHE warnings)
Drug misuse	Diagnosis of drug use disorder (lifetime: before or during incarceration—ICD-8: 304; ICD-9: 292, 304, 305 excl. 305A; ICD-10: F11-F19)	Warning type is drugs (taken from Niche warnings)
Any mental disorder	Diagnosis of any mental disorder excluding substance use disorders (lifetime: before or during incarceration)	Warning type is mental disorder (taken from Niche warnings)
Any severe mental disorder	ICD diagnosis of schizophrenia-spectrum or bipolar disorder (lifetime: before or during incarceration)	Not included

Note. Data on predictor variables in England were extracted from the PINs and the NICHE Record Management System. IMD = index of multiple deprivation; LSOA = lower-layer super output; PINs = police information notices

Table C.2: Summary of outcomes in the English sample and other OxRec samples

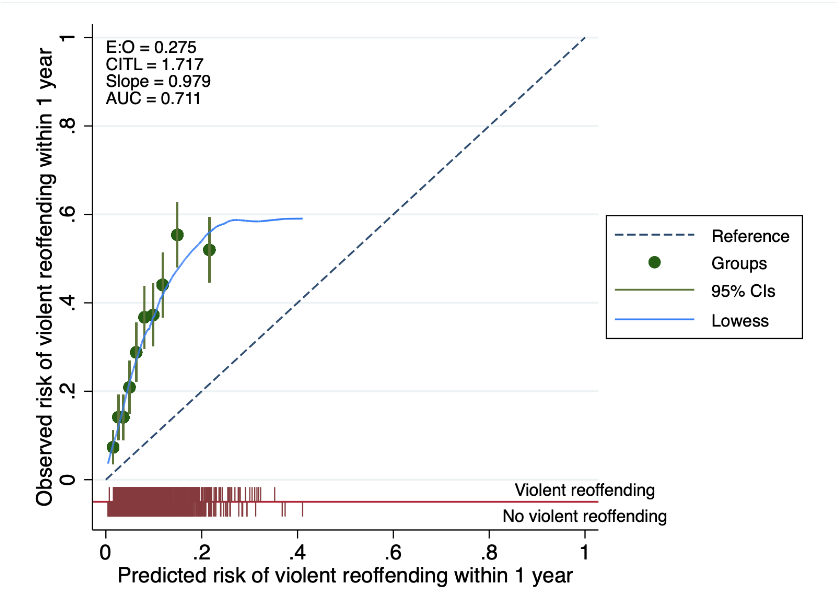
Outcomes	**England** n **= 1,770**	**Tajikistan** n **= 970**	**Netherlands** n **= 9,072**	**Sweden** n **= 37,100**
1 year violent reoffending	31%	15%	8%	12%
2 year violent reoffending	43%	*N/A*	16%	21%

Table C.3: Calibration performance measures for the OxRec model

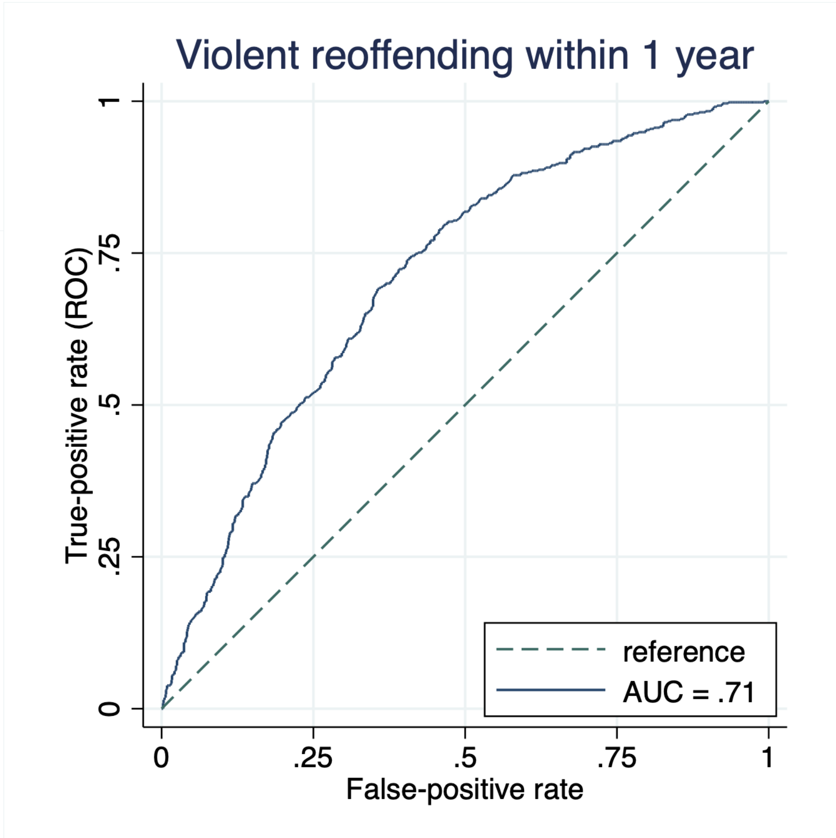
	Observed number of events	**Expected number of events (uncalibrated)**	**Ratio, Expected: Observed (uncalibrated) (95% CI)**	**Ratio of crude event rates, Sweden: England**	**Ratio, Expected: Observed (recalibrated) (95% CI)**	**Brier score**
1 year violent reoffending	550	152	0.28	0.39	1.01	0.19
2 year violent reoffending	765	251	0.33	0.49	1.01	0.22

Table C.4: Risk factors included in the final recalibrated model and their hazard ratios

Variable	**Multivariable adjusted hazard ratio (95% CI)**	
	England	**Sweden**
Sex (female)	0.75 (0.53–1.07)	0.51 (0.45–0.57)
Age	0.99 (0.98–1.00)	0.84 (0.83–0.85)
Immigrant	0.74 (0.48–1.14)	0.97 (0.92–1.02)
Length of incarceration		
$<$ 6 months	1 [Reference]	1 [Reference]
6-12 months	0.87 (0.68–1.11)	0.85 (0.81–0.90)
12–24 months	0.78 (0.61–1.01)	0.69 (0.63–0.75)
$\geq$ 24 months	0.62 (0.49–0.78)	0.55 (0.48–0.64)
Violent index offence	1.08 (0.90–1.30)	1.53 (1.46–1.59)
Previous violent crime	2.37 (1.88–2.99)	2.41 (2.29–2.54)
Civil status (unmarried)	0.87 (0.69–1.11)	1.08 (1.02–1.15)
Highest education		
$<$ 9 years		1 [Reference]
9–11 years		0.83 (0.79–0.87)
$\geq$ 12 years		0.65 (0.57–0.75)
Employment	0.94 (0.77–1.15)	0.68 (0.63–0.72)
Disposable income		
Negative		1 [Reference]
Zero		1.69 (1.11–2.57)
Low		1.45 (0.96–2.19)
Medium		1.45 (0.96–2.19)
High		1.57 (0.92–2.67)
Neighbourhood deprivation	1.10 (0.97–1.24)	1.03 (1.01–1.04)
Alcohol use disorder	1.41 (1.13–1.75)	1.41 (1.33–1.49)
Drug use disorder	1.30 (1.09–1.55)	1.51 (1.44–1.59)
Any mental disorder	1.08 (0.90–1.29)	1.09 (1.03-1.15)
Any severe mental disorder		1.10 (0.99–1.22)

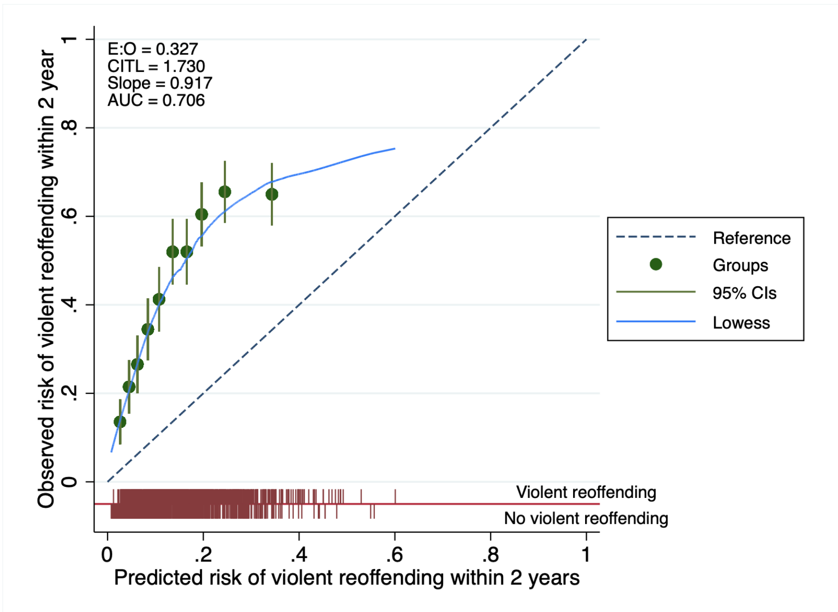


(a) Calibration plot

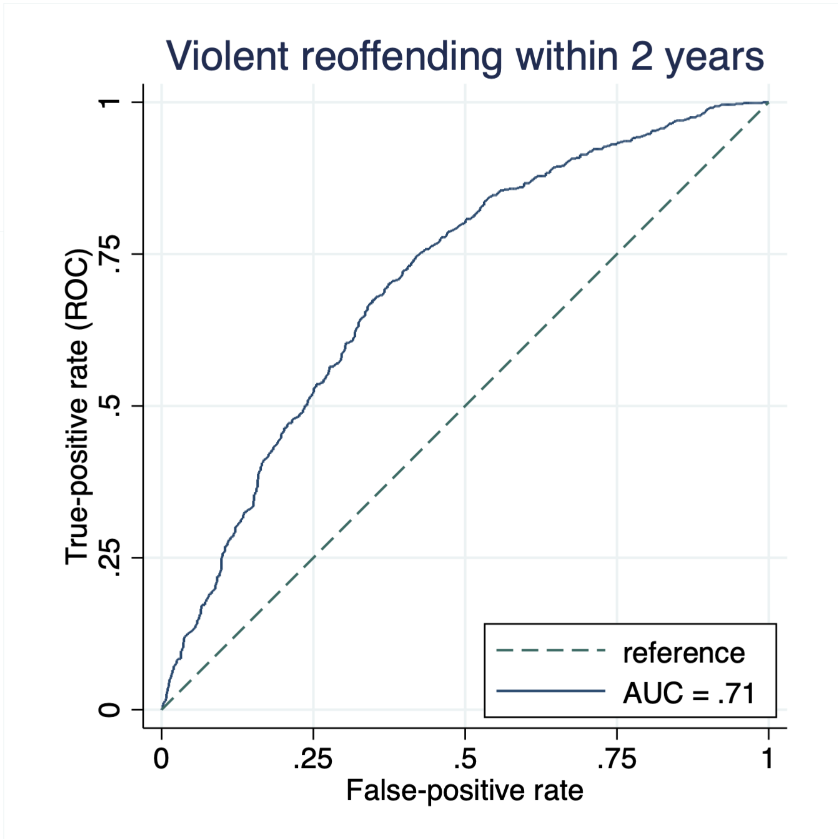


(b) ROC curve

Figure C.1: Model performance after simple validation (for violent reoffending within 1 year). AUC = area under the curve; CITL = calibration-in-the-large; E:O = expected to observed ratio; ROC = receiver operating characteristic.



(a) Calibration plot



(b) ROC curve

Figure C.2: Model performance after simple validation (for violent reoffending within 2 years). AUC = area under the curve; CITL = calibration-in-the-large; E:O = expected to observed ratio; ROC = receiver operating characteristic.

D
Appendix D: Psychological interventions in prison to reduce reoffending

D.1 Search strategy

Medline (Ovid MEDLINE® Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE®) 1946 to present

1. (Randomized Controlled Trial or Controlled Clinical Trial or Pragmatic Clinical Trial or Equivalence Trial or Clinical Trial, Phase III).pt. (615901)
2. Randomized Controlled Trial/ (522962)
3. exp Randomized Controlled Trials as Topic/ (143830)
4. "Randomized Controlled Trial (topic)"/ (0)
5. Controlled Clinical Trial/ (94068)
6. exp Controlled Clinical Trials as Topic/ (149298)
7. "Controlled Clinical Trial (topic)"/ (0)
8. Randomization/ (104661)
9. Random Allocation/ (104661)
10. Double-Blind Method/ (162383)
11. Double Blind Procedure/ (0)
12. Double-Blind Studies/ (162383)
13. Single-Blind Method/ (29726)
14. Single Blind Procedure/ (0)
15. Single-Blind Studies/ (29726)
16. Placebos/ (35331)
17. Placebo/ (0)
18. Control Groups/ (1718)
19. Control Group/ (1718)
20. (random* or sham or placebo*).ti,ab,hw,kw. (1570431)
21. ((singl* or doubl*) adj (blind* or dumm* or mask*)).ti,ab,hw,kw. (244156)
22. ((tripl* or trebl*) adj (blind* or dumm* or mask*)).ti,ab,hw,kw. (1137)
23. (control* adj3 (study or studies or trial* or group*)).ti,ab,kw. (1034249)
24. (Nonrandom* or non random* or non-random* or quasi-random* or quasirandom*).ti,ab,hw,kw. (46484)
25. allocated.ti,ab,hw. (69884)
26. ((open label or open-label) adj5 (study or studies or trial*)).ti,ab,hw,kw. (36928)
27. ((equivalence or superiority or non-inferiority or noninferiority) adj3 (study or studies or trial*)).ti,ab,hw,kw. (9184)
28. (pragmatic study or pragmatic studies).ti,ab,hw,kw. (441)
29. ((pragmatic or practical) adj3 trial*).ti,ab,hw,kw. (5689)
30. ((quasiexperimental or quasi-experimental) adj3 (study or studies or trial*)).ti,ab,hw,kw. (8685)
31. (phase adj3 (III or "3") adj3 (study or studies or trial*)).ti,hw,kw. (29714)
32. exp clinical study/ (982445)
33. (program* or intervention* or treatment* or therap* or trial*).ti,ab. (7868637)
34. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 (9014861)
35. prison/ (9972)
36. (prison* or incarcerat* or custod* or imprison* or detain* or inmate* or jail* or gaol* or penal* or penitentiary or "correctional facilit*" or probation*).ti,ab. (55114)
37. 35 or 36 (58191)
38. recidivism/ (368)

39. (recommit* or re-commit* or reoffend* or re-offend* or recidiv* or "repeat offen*" or reimprison* or re-imprison* or reincarcerat* or re-incarcerat* or reconvict* or re-convict* or rearrest* or re-arrest*).ti,ab. (7036)
40. 38 or 39 (7086)
41. 34 and 37 and 40 (1167)
42. exp experimental design/ (454407)
43. exp prisons/ (10539)
44. 36 or 43 (58640)
45. 34 or 42 (9106981)
46. 40 and 44 and 45 (1172)

Embase 1974 to present

1. (Randomized Controlled Trial or Controlled Clinical Trial or Pragmatic Clinical Trial or Equivalence Trial or Clinical Trial, Phase III).pt. (0)
2. Randomized Controlled Trial/ (649240)
3. exp Randomized Controlled Trials as Topic/ (197072)
4. "Randomized Controlled Trial (topic)"/ (197072)
5. Controlled Clinical Trial/ (469590)
6. exp Controlled Clinical Trials as Topic/ (204974)
7. "Controlled Clinical Trial (topic)"/ (11472)
8. Randomization/ (90411)
9. Random Allocation/ (86596)
10. Double-Blind Method/ (158182)
11. Double Blind Procedure/ (182720)
12. Double-Blind Studies/ (141482)
13. Single-Blind Method/ (40029)
14. Single Blind Procedure/ (42056)
15. Single-Blind Studies/ (42056)
16. Placebos/ (308772)
17. Placebo/ (364888)
18. Control Groups/ (110544)
19. Control Group/ (110544)
20. (random* or sham or placebo*).ti,ab,hw,kw. (2164252)
21. ((singl* or doubl*) adj (blind* or dumm* or mask*)).ti,ab,hw,kw. (319913)
22. ((tripl* or trebl*) adj (blind* or dumm* or mask*)).ti,ab,hw,kw. (1502)
23. (control* adj3 (study or studies or trial* or group*)).ti,ab,kw. (1450043)
24. (Nonrandom* or non random* or non-random* or quasi-random* or quasirandom*).ti,ab,hw,kw. (58495)
25. allocated.ti,ab,hw. (90765)
26. ((open label or open-label) adj5 (study or studies or trial*)).ti,ab,hw,kw. (68177)
27. ((equivalence or superiority or non-inferiority or noninferiority) adj3 (study or studies or trial*)).ti,ab,hw,kw. (13712)
28. (pragmatic study or pragmatic studies).ti,ab,hw,kw. (660)
29. ((pragmatic or practical) adj3 trial*).ti,ab,hw,kw. (6136)
30. ((quasiexperimental or quasi-experimental) adj3 (study or studies or trial*)).ti,ab,hw,kw. (13882)
31. (phase adj3 (III or "3") adj3 (study or studies or trial*)).ti,hw,kw. (98236)
32. exp clinical study/ (10346360)

33. (program* or intervention* or treatment* or therap* or trial*).ti,ab. (10542184)
34. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 (17192461)
35. prison/ (15593)
36. (prison* or incarcerat* or custod* or imprison* or detain* or inmate* or jail* or gaol* or penal* or penitentiary or "correctional facilit*" or probation*).ti,ab. (65520)
37. 35 or 36 (69306)
38. recidivism/ (3773)
39. (recommit* or re-commit* or reoffend* or re-offend* or recidiv* or "repeat offen*" or reimprison* or re-imprison* or reincarcerat* or re-incarcerat* or reconvict* or re-convict* or rearrest* or re-arrest*).ti,ab. (8948)
40. 38 or 39 (9746)
41. 34 and 37 and 40 (1787)

PsycINFO 1806 to present

1. (Randomized Controlled Trial or Controlled Clinical Trial or Pragmatic Clinical Trial or Equivalence Trial or Clinical Trial, Phase III).pt. (0)
2. Randomized Controlled Trial/ (657)
3. exp Randomized Controlled Trials as Topic/ (0)
4. "Randomized Controlled Trial (topic)"/ (0)
5. Controlled Clinical Trial/ (0)
6. exp Controlled Clinical Trials as Topic/ (0)
7. "Controlled Clinical Trial (topic)"/ (0)
8. Randomization/ (0)
9. Random Allocation/ (0)
10. Double-Blind Method/ (0)
11. Double Blind Procedure/ (0)
12. Double-Blind Studies/ (0)
13. Single-Blind Method/ (0)
14. Single Blind Procedure/ (0)
15. Single-Blind Studies/ (0)
16. Placebos/ (0)
17. Placebo/ (5907)
18. Control Groups/ (925)
19. Control Group/ (925)
20. (random* or sham or placebo*).ti,ab,hw,kw. (0)
21. ((singl* or doubl*) adj (blind* or dumm* or mask*)).ti,ab,hw,kw. (0)
22. ((tripl* or trebl*) adj (blind* or dumm* or mask*)).ti,ab,hw,kw. (0)
23. (control* adj3 (study or studies or trial* or group*)).ti,ab,kw. (0)
24. (Nonrandom* or non random* or non-random* or quasi-random* or quasirandom*).ti,ab,hw,kw. (0)
25. allocated.ti,ab,hw. (11111)
26. ((open label or open-label) adj5 (study or studies or trial*)).ti,ab,hw,kw.] (0)
27. ((equivalence or superiority or non-inferiority or noninferiority) adj3 (study or studies or trial*)).ti,ab,hw,kw. (0)
28. (pragmatic study or pragmatic studies).ti,ab,hw,kw. (0)
29. ((pragmatic or practical) adj3 trial*).ti,ab,hw,kw. (0)

30. ((quasiexperimental or quasi-experimental) adj3 (study or studies or trial*)).ti,ab,hw,kw. (0)
31. (phase adj3 (III or "3") adj3 (study or studies or trial*)).ti,hw,kw. (0)
32. exp clinical study/ (0)
33. (program* or intervention* or treatment* or therap* or trial*).ti,ab. (1438553)
34. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 (1444967)
35. prison/ (0)
36. (prison* or incarcerat* or custod* or imprison* or detain* or inmate* or jail* or gaol* or penal* or penitentiary or "correctional facilit*" or probation*).ti,ab. (57197)
37. 35 or 36 (57197)
38. recidivism/ (6122)
39. (recommit* or re-commit* or reoffend* or re-offend* or recidiv* or "repeat offen*" or reimprison* or re-imprison* or reincarcerat* or re-incarcerat* or reconvict* or re-convict* or rearrest* or re-arrest*).ti,ab. (10723)
40. 38 or 39 (11552)
41. 34 and 37 and 40 (2918)
42. exp experimental design/ (58966)
43. exp prisons/ (7320)
44. 36 or 43 (57440)
45. 34 or 42 (1477872)
46. 40 and 44 and 45 (2976)

Global Health <1973 to 2021 Week 06>

1. (Randomized Controlled Trial or Controlled Clinical Trial or Pragmatic Clinical Trial or Equivalence Trial or Clinical Trial, Phase III).pt. (0)
2. Randomized Controlled Trial/ (43800)
3. exp Randomized Controlled Trials as Topic/ (0)
4. "Randomized Controlled Trial (topic)"/ (0)
5. Controlled Clinical Trial/ (0)
6. exp Controlled Clinical Trials as Topic/ (0)
7. "Controlled Clinical Trial (topic)"/ (0)
8. Randomization/ (0)
9. Random Allocation/ (0)
10. Double-Blind Method/ (0)
11. Double Blind Procedure/ (0)
12. Double-Blind Studies/ (0)
13. Single-Blind Method/ (0)
14. Single Blind Procedure/ (0)
15. Single-Blind Studies/ (0)
16. Placebos/ (1927)
17. Placebo/ (1927)
18. Control Groups/ (0)
19. Control Group/ (0)
20. (random* or sham or placebo*).ti,ab,hw,kw. (0)
21. ((singl* or doubl*) adj (blind* or dumm* or mask*)).ti,ab,hw,kw. (0)
22. ((tripl* or trebl*) adj (blind* or dumm* or mask*)).ti,ab,hw,kw. (0)

23. (control* adj3 (study or studies or trial* or group*)).ti,ab,kw. (0)
24. (Nonrandom* or non random* or non-random* or quasi-random* or quasirandom*).ti,ab,hw,kw. (0)
25. allocated.ti,ab,hw. (12065)
26. ((open label or open-label) adj5 (study or studies or trial*)).ti,ab,hw,kw. (0)
27. ((equivalence or superiority or non-inferiority or noninferiority) adj3 (study or studies or trial*)).ti,ab,hw,kw. (0)
28. (pragmatic study or pragmatic studies).ti,ab,hw,kw. (0)
29. ((pragmatic or practical) adj3 trial*).ti,ab,hw,kw. (0)
30. ((quasiexperimental or quasi-experimental) adj3 (study or studies or trial*)).ti,ab,hw,kw. (0)
31. (phase adj3 (III or "3") adj3 (study or studies or trial*)).ti,hw,kw.] (0)
32. exp clinical study/ (0)
33. (program* or intervention* or treatment* or therap* or trial*).ti,ab. (1138544)
34. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 (1144983)
35. prison/ (0)
36. (prison* or incarcerat* or custod* or imprison* or detain* or inmate* or jail* or gaol* or penal* or penitentiary or "correctional facilit*" or probation*).ti,ab. (8586)
37. 35 or 36 (8586)
38. recidivism/ (0)
39. (recommit* or re-commit* or reoffend* or re-offend* or recidiv* or "repeat offen*" or reimprison* or re-imprison* or reincarcerat* or re-incarcerat* or reconvict* or re-convict* or rearrest* or re-arrest*).ti,ab. (474)
40. 38 or 39 (474)
41. 34 and 37 and 40 (114)
42. exp experimental design/ (1707)
43. exp prisons/ (2523)
44. 36 or 43 (8690)
45. 34 or 42 (1145956)
46. 40 and 44 and 45 (115)

Cochrane Database of Systematic Reviews
Cochrane Central Register of Controlled Trials

1. MeSH descriptor: [Prisons] explode all trees (126)
2. (prison* or incarcerat* or custod* or imprison* or detain* or inmate* or jail* or gaol* or penal* or penitentiary or "correctional facilit*" or probation*) (3538)
3. 1 or 2 (3539)
4. MeSH descriptor: [Recidivism] explode all trees (20)
5. (recommit* or re-commit* or reoffend* or re-offend* or recidiv* or "repeat offen*" or reimprison* or re-imprison* or reincarcerat* or re-incarcerat* or reconvict* or re-convict* or rearrest* or re-arrest*) (1608)
6. 4 or 5 (1608)
7. 3 and 6 (295)

(41 Cochrane Reviews; 250 Trials)

Google Scholar: I searched Google Scholar for articles which had cited the included studies. However, no additional study was identified.

Table D.1: Search results (February 17, 2021)

Search database	**Number of search results**
Ovid Medline	1172
Ovid Embase	1787
Ovid PsycINFO	2976
Ovid Global Health	115
Cochrane Database of Systematic Reviews and Cochrane CENTRAL	295
Google Scholar	0
Total	6345
Total (after deduplication)	4168

Table D.2: Ethnicity data for each study

Study	**Asian**	**Black/African American**	**Caucasian**	**Hispanic/Latin**	**Indigenous**	**Other**
Persons, 1967		40%	80%			
Annis, 1979			89%			11%
Lewis, 1983						
Linden and Perry, 1984						
Homant, 1986						
Shivrattan, 1988						
Lattimore et al., 1990			53%			47%
Guerra et al., 1990			40%			60%
Leeman et al., 1993		32%	67%	2%		
Robinson, 1995					12%	
Lindforss and Magnusson, 1997						
Dugan and Everett, 1998		2%	72%	22%	4%	
Ortmann, 2000						
Armstrong, 2003	7%	55%	32%	6%		
Prendergast et al., 2004		30%	38%	22%		10%
S. Sacks et al., 2004		30%	49%	17%		4%
Shapland et al., 2008						
Zlotnick et al., 2009		33%	47%	14%		6%
Messina et al., 2010		17%	48%	26%		9%
Proctor et al., 2012		24%	73%			3%
J. Y. Sacks et al., 2012			47%	26%		
Bowes et al., 2014	2%	2%	93%			3%
Yokotani and Tamura, 2015						
Kubiak et al., 2016		54%	46%			
Chaple et al., 2016		22%	49%	13%		17%
Malouf et al., 2017		48%	27%	15%		10%
Burraston and Eddy, 2017		13%	59%	8%	8%	11%
Hein et al., 2020		48%	17%	33%		2%
Gold et al., 2021						

Note. This table is based on the ethnicity data presented in each individual study, some of which only reported on specific ethnicities, and thus not all studies amount to 100%.

Table D.3: Quality assessment for each included study

Study	D1	D2	D3	D4	D5	Overall	Weight
Persons, 1967	Some concerns	Some concerns	Low	Low	Low	Some concerns	3.2
Annis, 1979	Some concerns	Some concerns	Low	Low	Low	Some concerns	3.6
Lewis, 1983	High	Some concerns	Low	Low	Low	High risk of bias	3.3
Linden and Perry, 1984	High	Some concerns	Low	Low	High	High risk of bias	2.5
Homant, 1986	Some concerns	Some concerns	Low	Low	High	High risk of bias	2.8
Shivrattan, 1988	Some concerns	Some concerns	Low	Low	Low	Some concerns	2.3
Lattimore et al., 1990	Some concerns	Some concerns	Low	Low	Low	Some concerns	4.6
Guerra et al., 1990a	Some concerns	Some concerns	Low	Low	Low	Some concerns	2.1
Guerra et al., 1990b	Some concerns	Some concerns	Low	Low	Low	Some concerns	2.1
Leeman et al., 1993	Some concerns	Some concerns	Low	Low	Low	Some concerns	2.0
Robinson, 1995	Some concerns	High	Low	Low	Low	High risk of bias	5.3
Lindforss and Magnusson, 1997	Some concerns	Some concerns	Low	Low	Some concerns	High risk of bias	2.3
Dugan and Everett, 1998	Some concerns	Some concerns	Low	Low	Low	Some concerns	4.1
Ortmann, 2000	Some concerns	Some concerns	Low	Low	Some concerns	High risk of bias	4.4
Armstrong, 2003	Some concerns	High	High	Low	Some concerns	High risk of bias	4.4
Prendergast et al., 2004	High	Some concerns	Low	Low	Low	High risk of bias	4.9
S. Sacks et al., 2004	Some concerns	Some concerns	Low	Low	Low	Some concerns	1.8
Shapland et al., 2008	Some concerns	Some concerns	Low	Low	Low	Some concerns	3.4
Zlotnick et al., 2009	Some concerns	Low	Low	Low	Low	Some concerns	2.2

Table D.3: Quality assessment for each included study (continued)

Study	D1	D2	D3	D4	D5	Overall	Weight
Messina et al., 2010	Some concerns	Low	Low	Low	Low	Some concerns	3.7
Proctor et al., 2012	Low	Low	Low	Low	Low	Low risk of bias	4.3
J. Y. Sacks et al., 2012	Some concerns	Some concerns	Low	Low	Low	Some concerns	4.4
Bowes et al., 2014	Some concerns	Low	Low	Low	Low	Some concerns	3.6
Yokotani and Tamura, 2015	Some concerns	Low	Low	Low	Low	Some concerns	2.3
Kubiak et al., 2016	High	Some concerns	Low	Low	Low	High risk of bias	1.7
Chaple et al., 2016	Some concerns	Low	Low	Low	Low	Some concerns	4.9
Malouf et al., 2017	Low	High	Low	Low	Low	High risk of bias	1.7
Burraston and Eddy, 2017	Low	Low	Low	Low	Low	Low risk of bias	5.0
Hein et al., 2020	Low	Some concerns	Low	Low	Low	Some concerns	4.3
Gold et al., 2021	Some concerns	Some concerns	Low	Low	Low	Some concerns	2.9

Note. D1 = Bias arising from the randomization process; D2 = Bias due to deviations from intended interventions; D3 = Bias due to missing outcome data; D4 = Bias in measurement of the outcome; D5 = Bias in selection of the reported result.

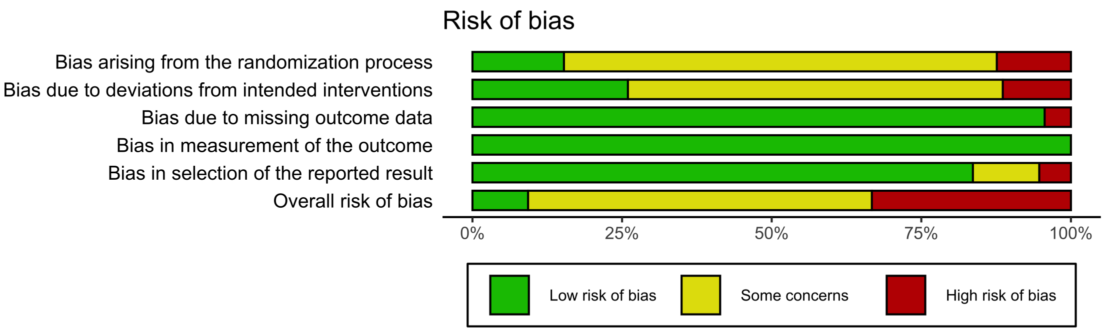


Figure D.1: Risk of bias summary plot

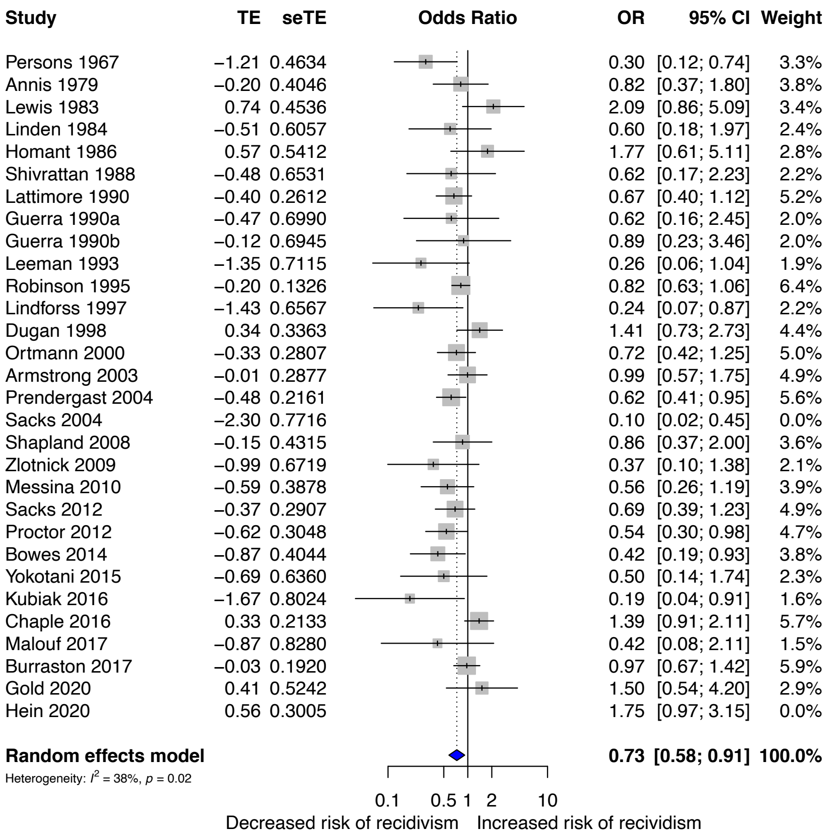


Figure D.2: Forest plot of all studies with two outliers removed (random-effects model). The two removed outliers are S. Sacks et al. (2004) and Hein et al. (2020).

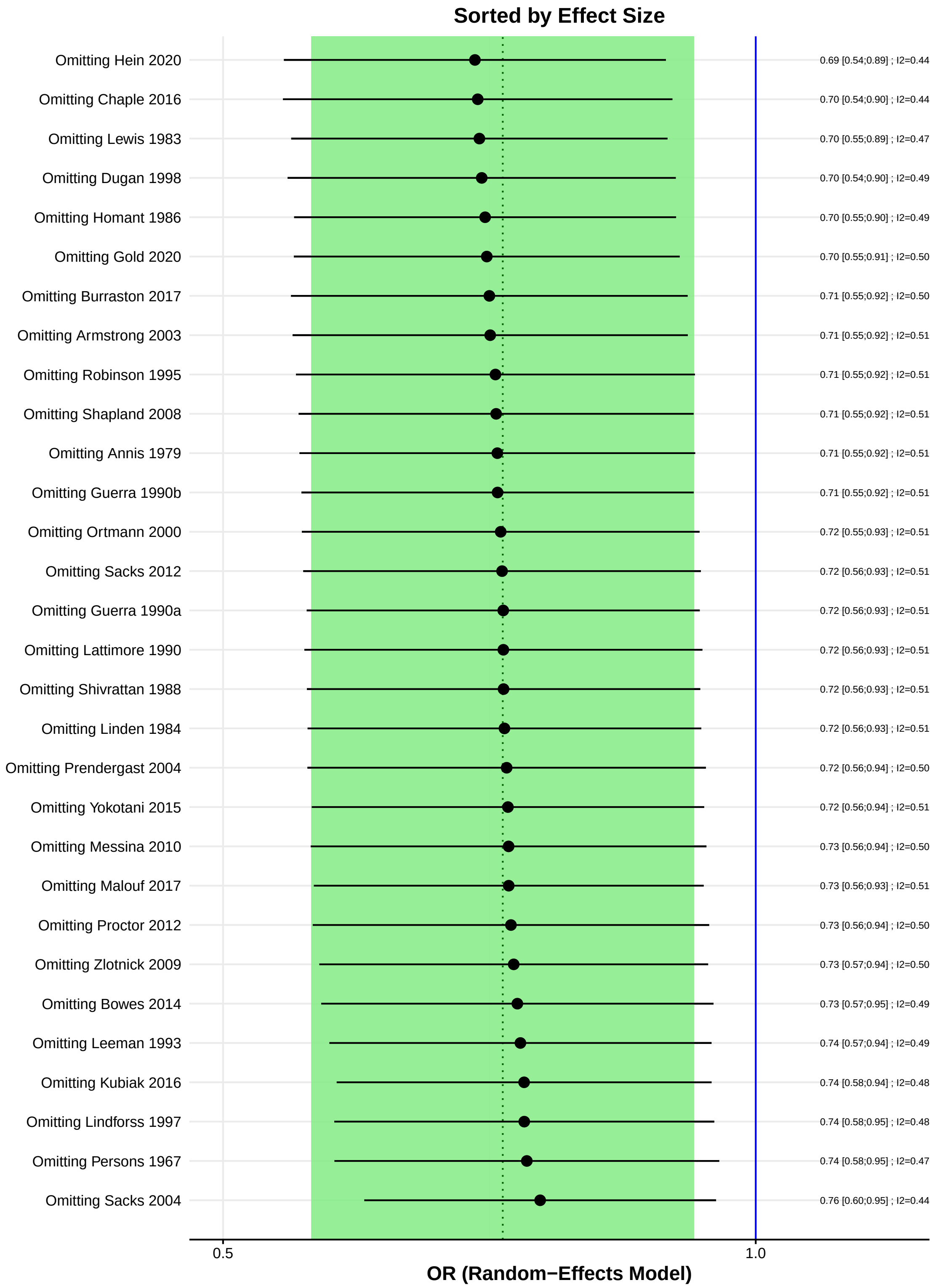


Figure D.3: Leave-One-Out analyses (sorted by effect size)

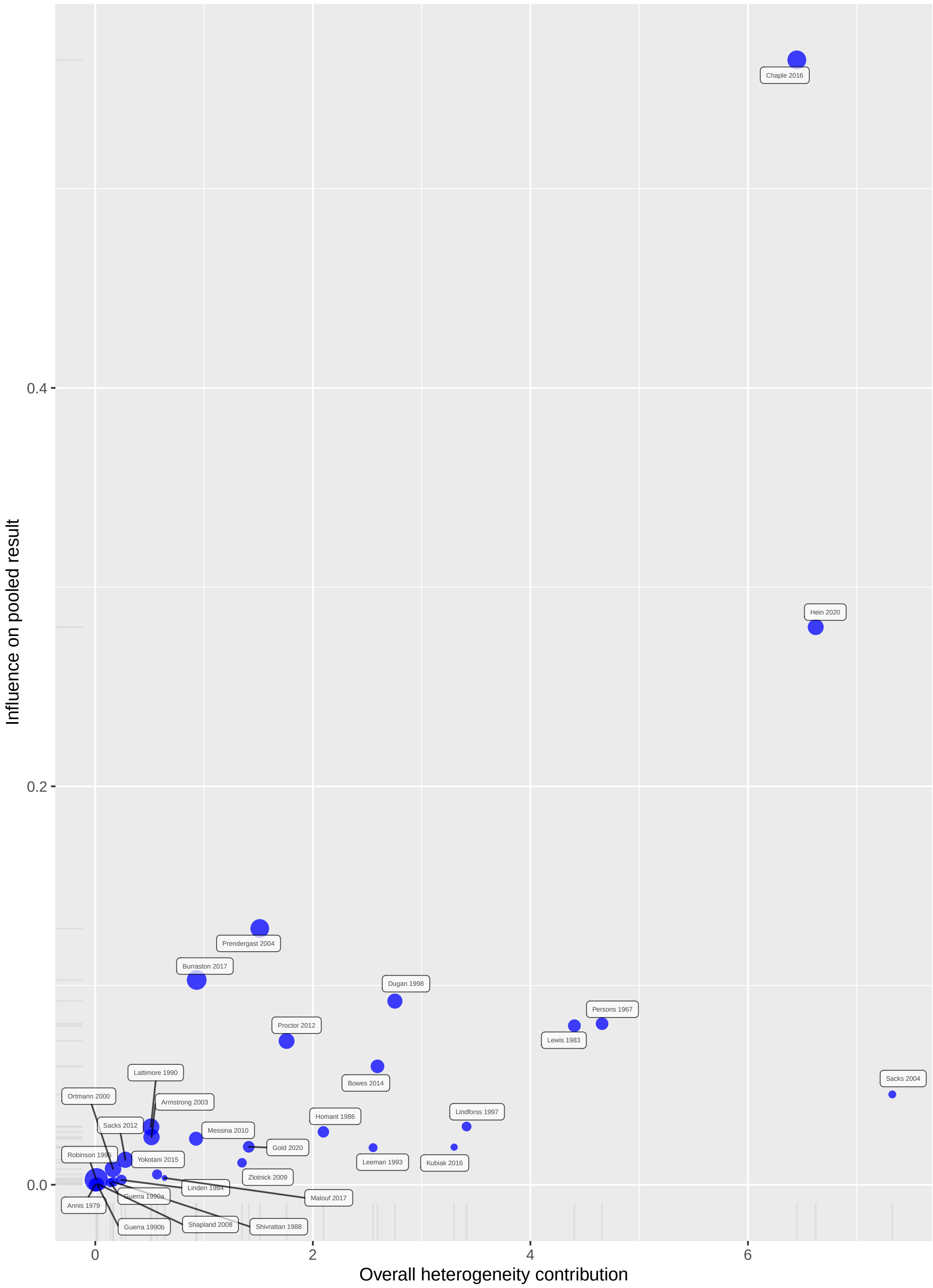


Figure D.4: Baujat plot

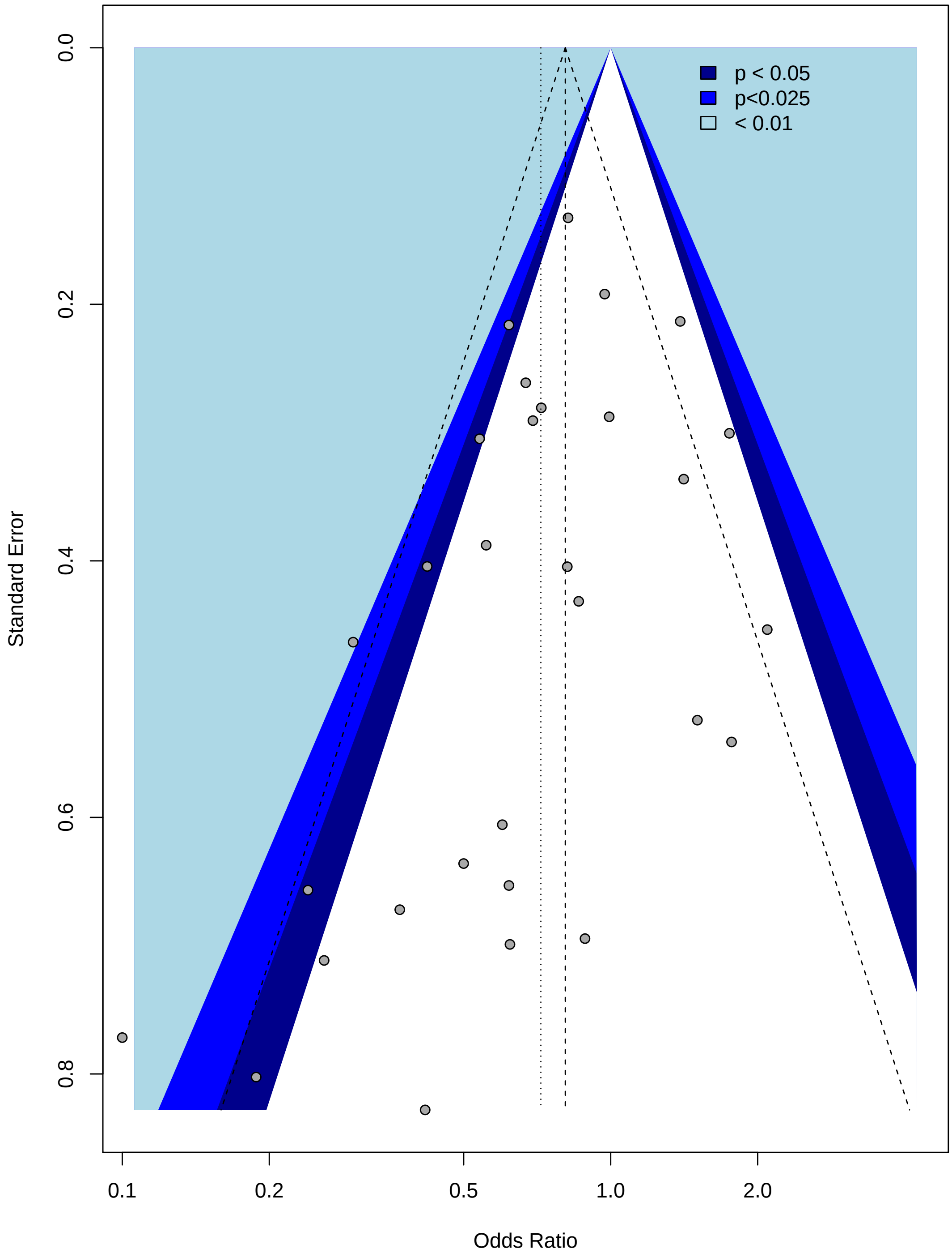


Figure D.5: Funnel plot for all studies

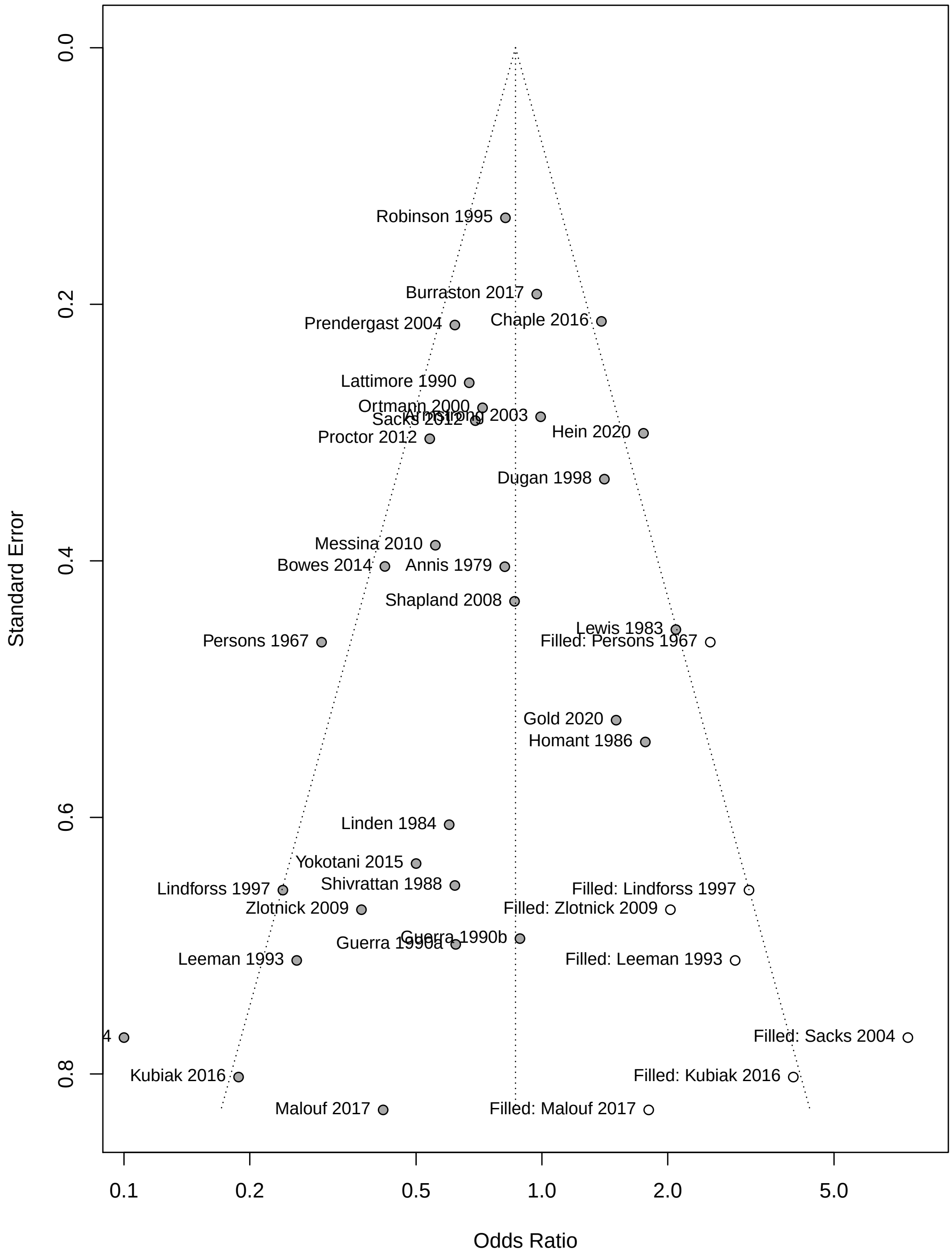


Figure D.6: Funnel plot for all studies (including imputed studies)

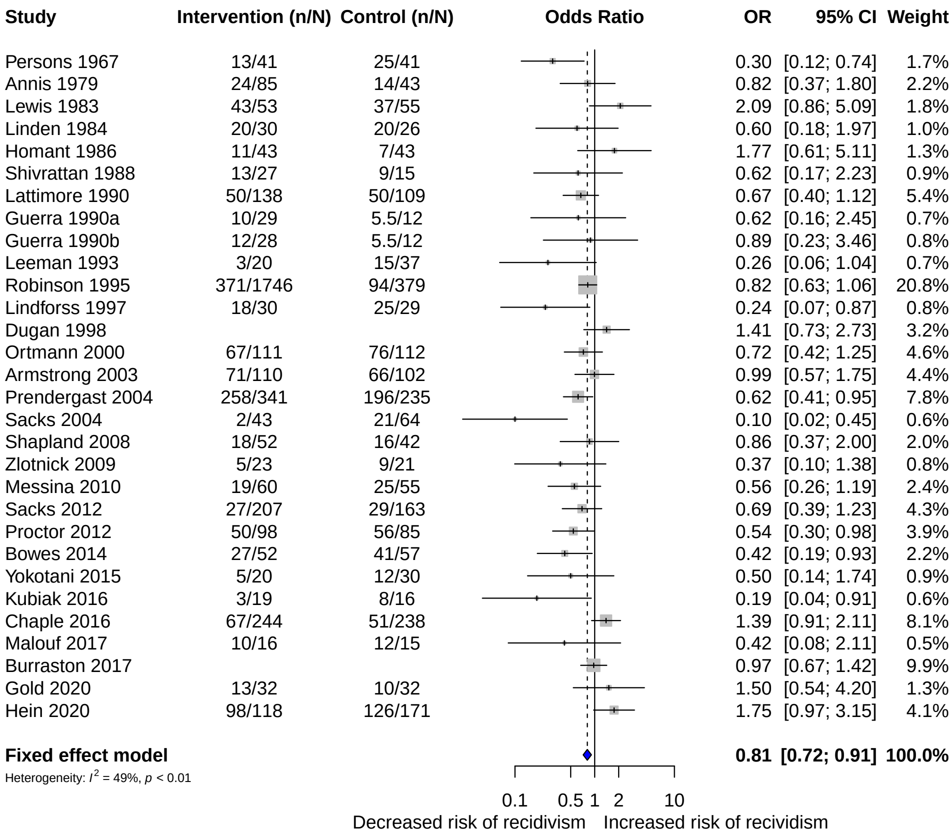


Figure D.7: Forest plot of all studies (fixed-effect model). Numbers of participants in the intervention and control groups are not available for Dugan and Everett (1998) and Burraston and Eddy (2017), as the outcome was presented as continuous data rather than dichotomous data in both of these studies.

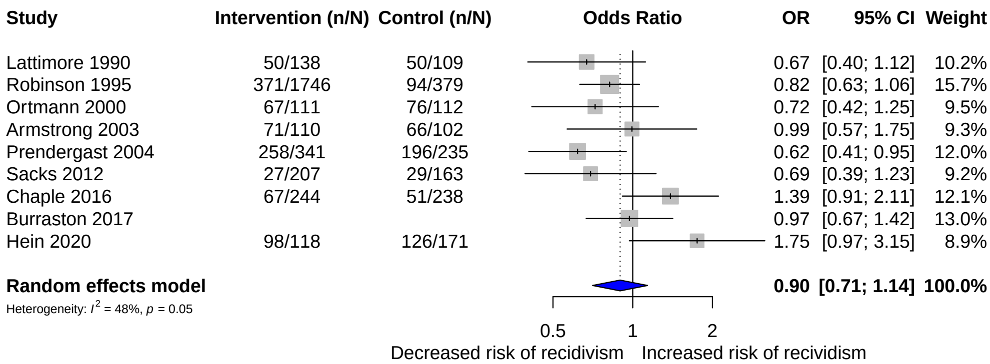


Figure D.8: Forest plot of studies with an intervention group of ≥100 participants (random-effects model). Numbers of participants in the intervention and control groups are not available for Burraston and Eddy (2017), as the outcome was presented as continuous data rather than dichotomous data in this study.

References

Abram, K. M., Teplin, L. A., Charles, D. R., Longworth, S. L., McClelland, G. M., & Dulcan, M. K. (2004). Posttraumatic stress disorder and trauma in youth in juvenile detention. *Archives of General Psychiatry*, *61*(4), 403–410.

Abrantes, A. M., Hoffmann, N. G., & Anton, R. (2005). Prevalence of co-occurring disorders among juveniles committed to detention centers. *International Journal of Offender Therapy and Comparative Criminology*, *49*(2), 179–193.

Adibi, A., Sadatsafavi, M., & Ioannidis, J. P. (2020). Validation and utility testing of clinical prediction models: Time to change the approach. *JAMA*, *324*(3), 235–236.

Aebi, M., Barra, S., Bessler, C., Steinhausen, H.-C., Walitza, S., & Plattner, B. (2016). Oppositional defiant disorder dimensions and subtypes among detained male adolescent offenders. *Journal of Child Psychology and Psychiatry*, *57*(6), 729–736.

Aebi, M., Linhart, S., Thun-Hohenstein, L., Bessler, C., Steinhausen, H.-C., & Plattner, B. (2015). Detained male adolescent offender's emotional, physical and sexual maltreatment profiles and their associations to psychiatric disorders and criminal behaviors. *Journal of Abnormal Child Psychology*, *43*(5), 999–1009.

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, *34*(3), 341–382.

Ahalt, C., Bolano, M., Wang, E. A., & Williams, B. (2015). The state of research funding from the National Institutes of Health for criminal justice health research. *Annals of Internal Medicine*, *162*(5), 345–352.

Aida, S. A., Aili, H. H., Manveen, K. S., Salwina, W. I. W., Subash, K. P., Ng, C. G., & Muhsin, A. Z. M. (2014). Prevalence of psychiatric disorders among juvenile offenders in Malaysian prisons and association with socio-demographic and personal factors. *International Journal of Prisoner Health*, *10*(2), 132–143.

Ako, T., Plugge, E., Mhlanga-Gunda, R., & Van Hout, M. (2020). Ethical guidance for health research in prisons in low-and middle-income countries: A scoping review. *Public Health*, *186*, 217–227.

Akobeng, A. K. (2005). Understanding randomised controlled trials. *Archives of Disease in Childhood*, *90*(8), 840–844.

Alper, M., Durose, M. R., & Markman, J. (2018). *2018 update on prisoner recidivism: A 9-year follow-up period (2005-2014)*. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. Washington, DC.

Altice, F. L., Azbel, L., Stone, J., Brooks-Pollock, E., Smyrnov, P., Dvoriak, S., Taxman, F. S., El-Bassel, N., Martin, N. K., Booth, R., et al. (2016). The perfect storm: Incarceration and the high-risk environment perpetuating transmission of HIV, hepatitis C virus, and tuberculosis in Eastern Europe and Central Asia. *Lancet*, *388*(10050), 1228–1248.

Anand, S., & Hanson, K. (1997). Disability-adjusted life years: A critical review. *Journal of Health Economics*, *16*(6), 685–702.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, *17*(1), 19–52.

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law*, *16*(1), 39–55.

Anjara, S. G., Bonetto, C., Van Bortel, T., & Brayne, C. (2020). Using the GHQ-12 to screen for mental health problems among primary care patients: Psychometrics and practical considerations. *International Journal of Mental Health Systems*, *14*(1), 1–13.

Annis, H. M. (1979). Group treatment of incarcerated offenders with alcohol and drug problems: A controlled evaluation. *Canadian Journal of Criminology*, *21*(1), 3–15.

Ardino, V. (2011). Post-traumatic stress in antisocial youth: A multifaceted reality. *Post-traumatic syndromes in childhood and adolescence* (pp. 211–230). John Wiley & Sons, Ltd.

Ardino, V. (2012). Offending behaviour: The role of trauma and PTSD. *European Journal of Psychotraumatology*, *3*(1), 18968.

Ardino, V., Milani, L., & Blasio, P. D. (2013). PTSD and re-offending risk: The mediating role of worry and a negative perception of other people's support. *European Journal of Psychotraumatology*, *4*(1), 21382.

Armstrong, T. A. (2003). The effect of moral reconation therapy on the recidivism of youthful offenders: A randomized experiment. *Criminal Justice and Behavior*, *30*(6), 668–687.

Asian Development Bank. (2021). *Basic statistics 2021.* Retrieved September 20, 2021, from https://www.adb.org/countries/tajikistan/poverty

Atilola, O. (2021). Mental disorders among detained youth: The hidden nature and peculiarities of African literature. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*(2), 202–203.

Atilola, O., Abiri, G., & Ola, B. (2020). Psychiatric morbidity among adolescents and youth involved with the juvenile justice system in sub-Saharan Africa: Systematic scoping review of current studies and research gaps. *International Journal of Law and Psychiatry*, *73*, 101633.

Atkins, D. L., Pumariega, A. J., Rogers, K., Montgomery, L., Nybro, C., Jeffers, G., & Sease, F. (1999). Mental health and incarcerated youth. i: Prevalence and nature of psychopathology. *Journal of Child and Family Studies*, *8*(2), 193–204.

Aubry, T., Nelson, G., & Tsemberis, S. (2015). Housing first for people with severe mental illness who are homeless: A review of the research and findings from the at home—chez soi demonstration project. *The Canadian Journal of Psychiatry*, *60*(11), 467–474.

Auty, K. M., Cope, A., & Liebling, A. (2017). Psychoeducational programs for reducing prison violence: A systematic review. *Aggression and Violent Behavior*, *33*, 126–143.

Avenevoli, S., Swendsen, J., He, J.-P., Burstein, M., & Merikangas, K. R. (2015). Major depression in the National Comorbidity Survey–Adolescent Supplement: Prevalence, correlates, and treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(1), 37–44.

Bachmann, S. (2018). Epidemiology of suicide and the psychiatric perspective. *International Journal of Environmental Research and Public Health*, *15*(7), 1425.

Baillargeon, J., Binswanger, I. A., Penn, J. V., Williams, B. A., & Murray, O. J. (2009). Psychiatric disorders and repeat incarcerations: The revolving prison door. *American Journal of Psychiatry*, *166*(1), 103–109.

Banerjee, A., Chen, S., Fatemifar, G., Zeina, M., Lumbers, R. T., Mielke, J., Gill, S., Kotecha, D., Freitag, D. F., Denaxas, S., et al. (2021). Machine learning for subtype definition and risk prediction in heart failure, acute coronary syndromes and atrial fibrillation: Systematic review of validity and clinical utility. *BMC Medicine*, *19*(1), 1–14.

Baranyi, G., Cassidy, M., Fazel, S., Priebe, S., & Mundt, A. P. (2018). Prevalence of posttraumatic stress disorder in prisoners. *Epidemiologic Reviews*, *40*(1), 134–145.

Baranyi, G., Scholl, C., Fazel, S., Patel, V., Priebe, S., & Mundt, A. P. (2019). Severe mental illness and substance use disorders in prisoners in low-income and middle-income countries: A systematic review and meta-analysis of prevalence studies. *Lancet Global Health*, *7*(4), e461–e471.

Barendregt, J. J., Doi, S. A., Lee, Y. Y., Norman, R. E., & Vos, T. (2013). Meta-analysis of prevalence. *Journal of Epidemiology and Community Health*, *67*(11), 974–978.

Baujat, B., Mahé, C., Pignon, J.-P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. *Statistics in Medicine*, *21*(18), 2641–2652.

Beaudry, G., Canal-Rivero, M., Ou, J., Matharu, J., Fazel, S., & Yu, R. (2022). Evaluating the risk of suicide and violence in severe mental illness: A feasibility study of two risk assessment tools (OxMIS and OxMIV) in general psychiatric settings. *Frontiers in Psychiatry*, 1297.

Beaudry, G., Zhong, S., Whiting, D., Javid, B., Frater, J., & Fazel, S. (2020). Managing outbreaks of highly contagious diseases in prisons: A systematic review. *BMJ Global Health*, *5*(11), e003201.

Bell, V. (2017). Open science in mental health research. *Lancet Psychiatry*, *4*(7), 525–526.

Bellass, S., Canvin, K., McLintock, K., & Wright, N. (2021). Integrating primary care across the prison and community interface. *British Journal of General Practice*, *71*(703), 56–57.

Bernardini, F., Attademo, L., Cleary, S. D., Luther, C., Shim, R. S., Quartesan, R., & Compton, M. T. (2017). Risk prediction models in psychiatry: Toward a new frontier for the prevention of mental illnesses. *The Journal of Clinical Psychiatry*, *78*(5), 18451.

Binswanger, I. A., Blatchford, P. J., Mueller, S. R., & Stern, M. F. (2013). Mortality after prison release: Opioid overdose and other causes of death, risk factors, and time trends from 1999 to 2009. *Annals of Internal Medicine*, *159*(9), 592–600.

Black, E. B., Ranmuthugala, G., Kondalsamy-Chennakesavan, S., Toombs, M. R., Nicholson, G. C., & Kisely, S. (2015). A systematic review: Identifying the prevalence rates of psychiatric disorder in Australia's indigenous populations. *Australian & New Zealand Journal of Psychiatry*, *49*(5), 412–429.

Bochenek, M. (2017). *World report 2016: Children behind bars: The global overuse of detention of children*. Retrieved May 3, 2022, from https://www.hrw.org/world-report/2016/country-chapters/africa-americas-asia-europe/central-asia-middle-east/north

Bolton, A. (1976). *A study of the need for and availability of mental health services for mentally disordered jail inmates and juveniles in detention facilities*. Arthur Bolton Associates. Sacramento, CA.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111.

Borenstein, M., & Higgins, J. (2013). Meta-analysis and subgroups. *Prevention Science*, *14*(2), 134–143.

Borges, G., Nock, M. K., Abad, J. M. H., Hwang, I., Sampson, N. A., Alonso, J., Andrade, L. H., Angermeyer, M. C., Beautrais, A., Bromet, E., et al. (2010). Twelve-month prevalence of and risk factors for suicide attempts in the World Health Organization World Mental Health Surveys. *The Journal of Clinical Psychiatry*, *71*(12), 1617–1628.

Borschmann, R., Janca, E., Carter, A., Willoughby, M., Hughes, N., Snow, K., Stockings, E., Hill, N. T., Hocking, J., Love, A., et al. (2020). The health of adolescents in detention: A global scoping review. *Lancet Public Health*, *5*(2), e114–e126.

Boutron, I., Dutton, S., Ravaud, P., & Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*, *303*(20), 2058–2064.

Bowes, N., McMurran, M., Evans, C., Oatley, G., Williams, B., & David, S. (2014). Treating alcohol-related violence: A feasibility study of a randomized controlled trial in prisons. *The Journal of Forensic Psychiatry & Psychology*, *25*(2), 152–163.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

Brooke, B. S., Schwartz, T. A., & Pawlik, T. M. (2021). MOOSE reporting guidelines for meta-analyses of observational studies. *JAMA Surgery*, *156*(8), 787–788.

Burns, M. E., Cook, S., Brown, L. M., Dague, L., Tyska, S., Romero, K. H., McNamara, C., & Westergaard, R. P. (2022). Association between assistance with Medicaid enrollment and use of health care after incarceration among adults with a history of substance use. *JAMA Network Open*, *5*(1), e2142688–e2142688.

Burraston, B. O., & Eddy, J. M. (2017). The moderating effect of living with a child before incarceration on postrelease outcomes related to a prison-based parent management training program. *Smith College Studies in Social Work*, *87*(1), 94–111.

Butchart, A., Mikton, C., & Krug, E. (2014). Governments must do more to address interpersonal violence. *Lancet*, *384*(9961), 2183–2185.

Byrne, M. W. (2005). Conducting research as a visiting scientist in a women's prison. *Journal of Professional Nursing*, *21*(4), 223–230.

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.

Campbell, M., Katikireddi, S. V., Sowden, A., & Thomson, H. (2019). Lack of transparency in reporting narrative synthesis of quantitative data: A methodological assessment of systematic reviews. *Journal of Clinical Epidemiology*, *105*, 1–9.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144.

Centers for Disease Control and Prevention. (2022). *Violence prevention at CDC.* Retrieved April 24, 2022, from https://www.cdc.gov/violenceprevention/about/index.html

Chandler, R. K., Fletcher, B. W., & Volkow, N. D. (2009). Treating drug abuse and addiction in the criminal justice system: Improving public health and safety. *JAMA*, *301*(2), 183–190.

Chang, Z., Lichtenstein, P., Långström, N., Larsson, H., & Fazel, S. (2016). Association between prescription of major psychotropic medications and violent reoffending after prison release. *JAMA*, *316*(17), 1798–1807.

Chang, Z., Lichtenstein, P., Larsson, H., & Fazel, S. (2015). Substance use disorders, psychiatric disorders, and mortality after release from prison: A nationwide longitudinal cohort study. *Lancet Psychiatry*, *2*(5), 422–430.

Chaple, M., Sacks, S., McKendrick, K., Marsch, L. A., Belenko, S., Leukefeld, C., Prendergast, M., & French, M. (2016). A comparative study of the therapeutic education system for incarcerated substance-abusing offenders. *The Prison Journal*, *96*(3), 485–508.

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., et al. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154–170.

Chen, J. H., & Asch, S. M. (2017). Machine learning and prediction in medicine—beyond the peak of inflated expectations. *New England Journal of Medicine*, *376*(26), 2507–2509.

Chiles, J. A., Miller, M. L., & Cox, G. B. (1980). Depression in an adolescent delinquent population. *Archives of General Psychiatry*, *37*(10), 1179–1184.

Chin, V., & Dandurand, Y. (2018). *Introductory handbook on the prevention of recidivism and the social reintegration of offenders.* United Nations Office on Drugs and Crime. Vienna, Austria.

Chitsabesan, P., Kroll, L., Bailey, S., Kenning, C., Sneider, S., MacDonald, W., & Theodosiou, L. (2006). Mental health needs of young offenders in custody and in the community. *The British Journal of Psychiatry*, *188*(6), 534–540.

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, *110*, 12–22.

Colins, O., Vermeiren, R., Schuyten, G., & Broekaert, E. (2009). Psychiatric disorders in property, violent, and versatile offending detained male adolescents. *American Journal of Orthopsychiatry*, *79*(1), 31–38.

Colins, O., Vermeiren, R., Vahl, P., Markus, M., Broekaert, E., & Doreleijers, T. (2011). Psychiatric disorder in detained male adolescents as risk factor for serious recidivism. *The Canadian Journal of Psychiatry*, *56*(1), 44–50.

Collins, G. S., de Groot, J. A., Dutton, S., Omar, O., Shanyinde, M., Tajar, A., Voysey, M., Wharton, R., Yu, L.-M., Moons, K. G., et al. (2014). External validation of multivariable prediction models: A systematic review of methodological conduct and reporting. *BMC Medical Research Methodology*, *14*(1), 1–11.

Collins, G. S., & Moons, K. G. (2019). Reporting of artificial intelligence prediction models. *Lancet*, *393*(10181), 1577–1579.

Collins, G. S., Ogundimu, E. O., & Altman, D. G. (2016). Sample size considerations for the external validation of a multivariable prognostic model: A resampling study. *Statistics in Medicine*, *35*(2), 214–226.

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Journal of British Surgery*, *102*(3), 148–158.

Collishaw, S., Maughan, B., Goodman, R., & Pickles, A. (2004). Time trends in adolescent mental health. *Journal of Child Psychology and Psychiatry*, *45*(8), 1350–1362.

Conroy, R. M., Pyörälä, K., Fitzgerald, A., Sans, S., Menotti, A., De Backer, G., De Bacquer, D., Ducimetiere, P., Jousilahti, P., Keil, U., et al. (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project. *European Heart Journal*, *24*(11), 987–1003.

Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, *115*(7), 928–935.

Copeland, W. E., Miller-Johnson, S., Keeler, G., Angold, A., & Costello, E. J. (2007). Childhood psychiatric disorders and young adult crime: A prospective, population-based study. *American Journal of Psychiatry*, *164*(11), 1668–1675.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797–806.

Cornish, R., Lewis, A., Parry, O. C., Ciobanasu, O., Mallett, S., & Fazel, S. (2019). A clinical feasibility study of the Forensic Psychiatry and Violence Oxford (FoVOx) tool. *Frontiers in Psychiatry*, *10*, 901.

Costello, E. J., Egger, H., & Angold, A. (2005). 10-year research update review: The epidemiology of child and adolescent psychiatric disorders: I. methods and public health burden. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(10), 972–986.

Cuellar, A. E., Markowitz, S., & Libby, A. M. (2004). Mental health and substance abuse treatment and juvenile crime. *Journal of Mental Health Policy and Economics*, *7*(2), 59–68.

Cuijpers, P., & Cristea, I.-A. (2016). How to prove that your therapy is effective, even when it is not: A guideline. *Epidemiology and Psychiatric Sciences*, *25*(5), 428–435.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, *40*(2), 211–223.

Cyranowski, J. M., Frank, E., Young, E., & Shear, M. K. (2000). Adolescent onset of the gender difference in lifetime rates of major depression: A theoretical model. *Archives of General Psychiatry*, *57*(1), 21–27.

Dahlberg, L. L., & Mercy, J. A. (2009). The history of violence as a public health issue. *AMA Virtual Mentor*, *11*(2), 167–172.

Das-Munshi, J., Ford, T., Hotopf, M., Prince, M., & Stewart, R. (2020). *Practical psychiatric epidemiology*. Oxford University Press.

De Andrade, D., Ritchie, J., Rowlands, M., Mann, E., & Hides, L. (2018). Substance use and recidivism outcomes for prison-based drug and alcohol interventions. *Epidemiologic Reviews*, *40*(1), 121–133.

Debidin, M. (2009). *A compendium of research and analysis on the offender assessment system (OASys) 2006-2009.* Ministry of Justice. London, UK.

Debray, T. P., Vergouwe, Y., Koffijberg, H., Nieboer, D., Steyerberg, E. W., & Moons, K. G. (2015). A new framework to enhance the interpretation of external validation studies of clinical prediction models. *Journal of Clinical Epidemiology*, *68*(3), 279–289.

Decker, M. R., Wilcox, H. C., Holliday, C. N., & Webster, D. W. (2018). An integrated public health approach to interpersonal violence and suicide prevention and response. *Public Health Reports*, *133*(1_suppl), 65S–79S.

de Jong, V. M. T., Rousset, R. Z., Antonio-Villa, N. E., Buenen, A. G., Van Calster, B., Bello-Chavolla, O. Y., Brunskill, N. J., Curcin, V., Damen, J. A. A., Fermín-Martínez, C. A., Fernández-Chirino, L., Ferrari, D., Free, R. C., Gupta, R. K., Haldar, P., Hedberg, P., Korang, S. K., Kurstjens, S., Kusters, R., . . . Debray, T. P. A. (2022). Clinical prediction models for mortality in patients with covid-19: External validation and individual participant data meta-analysis. *BMJ*, *378*, e069881.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188.

Derzon, J. H. (2001). Antisocial behavior and the prediction of violence: A meta-analysis. *Psychology in the Schools*, *38*(2), 93–106.

Desmarais, S. L., Johnson, K. L., & Singh, J. P. (2016). Performance of recidivism risk assessment instruments in us correctional settings. *Psychological Services*, *13*(3), 206–222.

Dhiman, P., Ma, J., Andaur Navarro, C. L., Speich, B., Bullock, G., Damen, J. A., Hooft, L., Kirtley, S., Riley, R. D., Van Calster, B., et al. (2022). Methodological conduct of prognostic prediction models developed using machine learning in oncology: A systematic review. *BMC Medical Research Methodology*, *22*(1), 1–16.

Dimond, C., & Misch, P. (2002). Psychiatric morbidity in children remanded to prison custody—a pilot study. *Journal of Adolescence*, *25*(6), 681–689.

Dixon, A., Howie, P., & Starling, J. (2004). Psychopathology in female juvenile offenders. *Journal of Child Psychology and Psychiatry*, *45*(6), 1150–1158.

Dória, G. M. S., Antoniuk, S. A., Assumpção Junior, F. B., Fajardo, D. N., & Ehlke, M. N. (2015). Delinquency and association with behavioral disorders and substance abuse. *Revista da Associação Médica Brasileira*, *61*(1), 51–57.

Douglas, K. S., & Reeves, K. A. (2010). Historical-Clinical-Risk Management-20 (HCR-20) Violence Risk Assessment Scheme: Rationale, application, and empirical overview. In R. K. Otto & K. S. Douglas (Eds.), *Injury prevention and environmental health. 3rd edition.* (pp. 147–185). Handbook of violence risk assessment.

Dowden, C., & Andrews, D. A. (1999). What works for female offenders: A meta-analytic review. *Crime & Delinquency*, *45*(4), 438–452.

Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending: A meta-analysis. *Canadian Journal of Criminology*, *42*(4), 449–467.

Duclos, C. W., Beals, J., Novins, D. K., Martin, C., Jewett, C. S., & Manson, S. M. (1998). Prevalence of common psychiatric disorders among American Indian adolescent detainees. *Journal of the American Academy of Child & Adolescent Psychiatry*, *37*(8), 866–873.

Dugan, J. R., & Everett, R. S. (1998). An experimental test of chemical dependency therapy for jail inmates. *International Journal of Offender Therapy and Comparative Criminology*, *42*(4), 360–368.

Dumont, D. M., Brockmann, B., Dickman, S., Alexander, N., & Rich, J. D. (2012). Public health and the epidemic of incarceration. *Annual Review of Public Health*, *33*, 325–339.

Durose, M. R., Cooper, A. D., & Snyder, H. N. (2014). *Recidivism of prisoners released in 30 states in 2005: Patterns from 2005 to 2010.* US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. Washington, DC.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463.

Dziuba-Leatherman, J., & Finkelhor, D. (1994). How does receiving information about sexual abuse influence boys' perceptions of their risk? *Child Abuse & Neglect*, *18*(7), 557–568.

Eaneff, S., Obermeyer, Z., & Butte, A. J. (2020). The case for algorithmic stewardship for artificial intelligence and machine learning technologies. *JAMA*, *324*(14), 1397–1398.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634.

Elkington, K. S., Teplin, L. A., Abram, K. M., Jakubowski, J. A., Dulcan, M. K., & Welty, L. J. (2015). Psychiatric disorders and violence: A study of delinquent youth after detention. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(4), 302–312.

Erskine, H. E., Ferrari, A. J., Nelson, P., Polanczyk, G. V., Flaxman, A. D., Vos, T., Whiteford, H. A., & Scott, J. G. (2013). Research review: Epidemiological modelling of attention-deficit/hyperactivity disorder and conduct disorder for the Global Burden of Disease Study 2010. *Journal of Child Psychology and Psychiatry*, *54*(12), 1263–1274.

Facer-Irwin, E., Blackwood, N. J., Bird, A., Dickson, H., McGlade, D., Alves-Costa, F., & MacManus, D. (2019). PTSD in prison settings: A systematic review and meta-analysis of comorbid mental disorders and problematic behaviours. *PLOS One*, *14*(9), e0222407.

Fair, H., & Walmsley, R. (2021). *World prison population list, thirteenth edition.* Institute for Crime & Justice Policy Research.

Fairchild, G., Hawes, D. J., Frick, P. J., Copeland, W. E., Odgers, C. L., Franke, B., Freitag, C. M., & De Brito, S. A. (2019). Conduct disorder. *Nature Reviews Disease Primers*, *5*(1), 1–25.

Farrell, A. (2000). Women, crime and drugs: Testing the effect of therapeutic communities. *Women & Criminal Justice*, *11*(1), 21–48.

Farrington, D. P., & Welsh, B. C. (2005). Randomized experiments in criminology: What have we learned in the last two decades? *Journal of Experimental Criminology*, *1*(1), 9–38.

Favril, L., Yu, R., Hawton, K., & Fazel, S. (2020). Risk factors for self-harm in prison: A systematic review and meta-analysis. *Lancet Psychiatry*, *7*(8), 682–691.

Fazel, S. (2019). The scientific validity of current approaches to violence and criminal risk assessment. In J. W. de Keijser, J. V. Roberts, & J. Ryberg (Eds.), *Predictive* *sentencing: Normative and empirical perspectives* (pp. 197–212). Bloomsbury Publishing.

Fazel, S., & Baillargeon, J. (2011). The health of prisoners. *Lancet*, *377*(9769), 956–965.

Fazel, S., Bains, P., & Doll, H. (2006). Substance abuse and dependence in prisoners: A systematic review. *Addiction*, *101*(2), 181–191.

Fazel, S., Bromberg, D. J., & Altice, F. L. (2022). HIV, substance use, and mental health care in prisons. *Lancet Psychiatry*.

Fazel, S., Burghart, M., Fanshawe, T., Gil, S. D., Monahan, J., & Yu, R. (2022). The predictive performance of criminal risk assessment tools used at sentencing: Systematic review of validation studies. *Journal of Criminal Justice*, *81*, 101902.

Fazel, S., Cartwright, J., Norman-Nott, A., & Hawton, K. (2008). Suicide in prisoners: A systematic review of risk factors. *The Journal of Clinical Psychiatry*, *69*(11), 1721–1731.

Fazel, S., Chang, Z., Fanshawe, T., Langstrom, N., Lichtenstein, P., Larsson, H., & Mallett, S. (2016). Prediction of violent reoffending on release from prison: Derivation and external validation of a scalable tool. *Lancet Psychiatry*, *3*(6), 535–543.

Fazel, S., & Danesh, J. (2002). Serious mental disorder in 23 000 prisoners: A systematic review of 62 surveys. *Lancet*, *359*(9306), 545–550.

Fazel, S., Doll, H., & Långström, N. (2008). Mental disorders among adolescents in juvenile detention and correctional facilities: A systematic review and metaregression analysis of 25 surveys. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(9), 1010–1019.

Fazel, S., Hayes, A. J., Bartellas, K., Clerici, M., & Trestman, R. (2016). Mental health of prisoners: Prevalence, adverse outcomes, and interventions. *Lancet Psychiatry*, *3*(9), 871–881.

Fazel, S., Ramesh, T., & Hawton, K. (2017). Suicide in prisons: An international study of prevalence and contributory factors. *Lancet Psychiatry*, *4*(12), 946–952.

Fazel, S., Sariaslan, A., & Fanshawe, T. (2022). Towards a more evidence-based risk assessment for people in the criminal justice system: The case of OxRec in the Netherlands. *European Journal on Criminal Policy and Research*, 1–10.

Fazel, S., & Seewald, K. (2012). Severe mental illness in 33 588 prisoners worldwide: Systematic review and meta-regression analysis. *The British Journal of Psychiatry*, *200*(5), 364–373.

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: Systematic review and meta-analysis. *BMJ*, *345*(e4692).

Fazel, S., Smith, E. N., Chang, Z., & Geddes, J. R. (2018). Risk factors for interpersonal violence: An umbrella review of meta-analyses. *The British Journal of Psychiatry*, *213*(4), 609–614.

Fazel, S., & Wolf, A. (2015). A systematic review of criminal recidivism rates worldwide: Current difficulties and recommendations for best practice. *PLOS One*, *10*(6), 1–8.

Fazel, S., Wolf, A., Larsson, H., Lichtenstein, P., Mallett, S., & Fanshawe, T. R. (2017). Identification of low risk of violent crime in severe mental illness with a clinical prediction tool (Oxford Mental Illness and Violence tool [OxMIV]): A derivation and validation study. *Lancet Psychiatry*, *4*(6), 461–468.

Fazel, S., Wolf, A., Vazquez-Montes, M. D., & Fanshawe, T. R. (2019). Prediction of violent reoffending in prisoners and individuals on probation: A Dutch validation study (OxRec). *Scientific Reports*, *9*(1), 1–9.

Fazel, S., Yoon, I. A., & Hayes, A. J. (2017). Substance use disorders in prisoners: An updated systematic review and meta-regression analysis in recently incarcerated men and women. *Addiction*, *112*(10), 1725–1739.

Federal Bureau of Investigation (US). (2020). *Crime in the United States 2020*. Retrieved April 29, 2022, from https://crime-data-explorer.app.cloud.gov/pages/explorer/crime/crime-trend

Florkowski, C. M. (2008). Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: Communicating the performance of diagnostic tests. *The Clinical Biochemist Reviews*, *29*(Suppl 1), S83–S87.

Fombonne, E. (1998). Increased rates of psychosocial disorders in youth. *European Archives of Psychiatry and Clinical Neuroscience*, *248*(1), 14–21.

Fontanarosa, J., Uhl, S., Oyesanmi, O., & Schoelles, K. M. (2013). *Interventions for adult offenders with serious mental illness*. Agency for Healthcare Research and Quality (US). Rockville, MD.

Forsman, J., Cornish, R., & Fazel, S. (2022). Integrating static and modifiable risk factors in violence risk assessment for forensic psychiatric patients: A feasibility study of FoVOx. *Nordic Journal of Psychiatry*, 1–7.

Fovet, T., Chan-Chee, C., Baillet, M., Horn, M., Wathelet, M., D'Hondt, F., Thomas, P., Amad, A., & Lamer, A. (2022). Psychiatric hospitalisations for people who are incarcerated, 2009–2019: An 11-year retrospective longitudinal study in France. *EClinicalMedicine*, *46*, 101374.

Fox, K., Zambrana, K., & Lane, J. (2011). Multivariate comparison of male and female adolescent substance abusers with accompanying legal problems. *Journal of Criminal Justice Education*, *22*(2), 304–327.

Frank, T. D., Carter, A., Jahagirdar, D., Biehl, M. H., Douwes-Schultz, D., Larson, S. L., Arora, M., Dwyer-Lindgren, L., Steuben, K. M., Abbastabar, H., et al. (2019). Global, regional, and national incidence, prevalence, and mortality of HIV, 1980–2017, and forecasts to 2030, for 195 countries and territories: a systematic analysis for the Global Burden of Diseases, Injuries, and Risk Factors Study 2017. *Lancet HIV*, *6*(12), e831–e859.

Fusar-Poli, P., Hijazi, Z., Stahl, D., & Steyerberg, E. W. (2018). The science of prognosis in psychiatry: A review. *JAMA Psychiatry*, *75*(12), 1289–1297.

Fusar-Poli, P., Rutigliano, G., Stahl, D., Davies, C., Bonoldi, I., Reilly, T., & McGuire, P. (2017). Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry*, *74*(5), 493–500.

Fusar-Poli, P., Stringer, D., M.S. Durieux, A., Rutigliano, G., Bonoldi, I., De Micheli, A., & Stahl, D. (2019). Clinical-learning versus machine-learning for transdiagnostic prediction of psychosis onset in individuals at-risk. *Translational Psychiatry*, *9*(1), 1–11.

Gaïffas, A., Galéra, C., Mandon, V., & Bouvard, M. P. (2014). Attention-deficit/hyperactivity disorder in young french male prisoners. *Journal of Forensic Sciences*, *59*(4), 1016–1019.

Ganna, A., & Ingelsson, E. (2015). 5 year mortality predictors in 498 103 UK Biobank participants: A prospective population-based study. *Lancet*, *386*(9993), 533–540.

Garbarino, J. (1985). *Adolescent development: An ecological perspective*. Charles E. Merrill Publishing Company.

Garbarino, J. (1995). *Raising children in a socially toxic environment*. ERIC.

Gersons, B. P., & Carlier, I. V. (1992). Post-traumatic stress disorder: The history of a recent concept. *The British Journal of Psychiatry*, *161*(6), 742–748.

Gerstein, H. C., McMurray, J., & Holman, R. R. (2019). Real-world studies no substitute for RCTs in establishing efficacy. *Lancet*, *393*(10168), 210–211.

Ghanizadeh, A., Nouri, S. Z., & Nabr, S. S. (2012). Psychiatric problems and suicidal behaviour in incarcerated adolescents in the Islamic Republic of Iran. *Eastern Mediterranean Health Journal*, *18*(4), 311–317.

Gold, C., Due, F. B., Thieu, E. K., Hjørnevik, K., Tuastad, L., & Assmus, J. (2021). Long-term effects of short-term music therapy for prison inmates: Six-year follow-up of a randomized controlled trial. *International Journal of Offender Therapy and Comparative Criminology*, *65*(5), 543–557.

Gonzalvo, G. O. (2002). Estado de salud y nutrición de mujeres adolescentes delincuentes [Health and nutritional status of delinquent female adolescents]. *Anales de Pediatría*, *56*(2), 116–120.

Gosden, N. P., Kramp, P., Gabrielsen, G., & Sestoft, D. (2003). Prevalence of mental disorders among 15–17-year-old male adolescent remand prisoners in Denmark. *Acta Psychiatrica Scandinavica*, *107*(2), 102–110.

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.

Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, *81*(3), 515–526.

Gretton, H. M., & Clift, R. J. (2011). The mental health needs of incarcerated youth in British Columbia, Canada. *International Journal of Law and Psychiatry*, *34*(2), 109–115.

Grove, L., King, C. M., Bomysoad, R., Vasquez, L., & Kois, L. E. (2021). Technology for assessment and treatment of justice-involved youth: A systematic literature review. *Law and Human Behavior*, *45*(5), 413.

Guchereau, M., Jourkiv, O., & Zametkin, A. J. (2009). Mental disorders among adolescents in juvenile detention and correctional facilities: Posttraumatic stress disorder is overlooked. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(3), 340.

Guebert, A. F., & Olver, M. E. (2014). An examination of criminogenic needs, mental health concerns, and recidivism in a sample of violent young offenders: Implications for risk, need, and responsivity. *International Journal of Forensic Mental Health*, *13*(4), 295–310.

Guerra, N. G., Huesmann, L. R., & Zelli, A. (1990). Attributions for social failure and aggression in incarcerated delinquent youth. *Journal of Abnormal Child Psychology*, *18*(4), 347–355.

Guinney, J., Wang, T., Laajala, T. D., Winner, K. K., Bare, J. C., Neto, E. C., Khan, S. A., Peddinti, G., Airola, A., Pahikkala, T., et al. (2017). Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: Development of a prognostic model through a crowdsourced challenge with open clinical trial data. *Lancet Oncology*, *18*(1), 132–142.

Gulati, G., Upshaw, J., Wessler, B. S., Brazil, R. J., Nelson, J., van Klaveren, D., Lundquist, C. M., Park, J. G., McGinnes, H., Steyerberg, E. W., et al. (2022).

Generalizability of cardiovascular disease clinical prediction models: 158 independent external validations of 104 unique models. *Circulation: Cardiovascular Quality and Outcomes*, *15*(4), e008487.

Gureje, O., & Abdulmalik, J. (2019). Severe mental disorders among prisoners in low-income and middle-income countries: Reaching the difficult to reach. *Lancet Global Health*, *7*(4), e392–e393.

Haglund, A., Tidemalm, D., Jokinen, J., Långström, N., Lichtenstein, P., Fazel, S., & Runeson, B. (2014). Suicide after release from prison: A population-based cohort study from Sweden. *The Journal of Clinical Psychiatry*, *75*(10), 20451.

Halligan, S., Altman, D. G., & Mallett, S. (2015). Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: A discussion and proposal for an alternative approach. *European Radiology*, *25*(4), 932–939.

Hamerlynck, S. M., Cohen-Kettenis, P. T., Vermeiren, R., Jansen, L., Bezemer, P. D., & Doreleijers, T. A. (2007). Sexual risk behavior and pregnancy in detained adolescent females: A study in Dutch detention centers. *Child and Adolescent Psychiatry and Mental Health*, *1*(1), 1–7.

Haniffa, R., Isaam, I., De Silva, A. P., Dondorp, A. M., & De Keizer, N. F. (2018). Performance of critical care prognostic scoring systems in low and middle-income countries: A systematic review. *Critical Care*, *22*(1), 1–22.

Hanson, R. K. (2017). Assessing the calibration of actuarial risk scales: A primer on the E/O index. *Criminal Justice and Behavior*, *44*(1), 26–39.

Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior*, *36*(9), 865–891.

Harcourt, B. E. (2006). *Against prediction: Profiling, policing, and punishing in an actuarial age*. University of Chicago Press.

Hardwicke, T. E., & Wagenmakers, E.-J. (2022). Reducing bias, increasing transparency, and calibrating confidence with preregistration. *MetaArXiv*.

Harel, O., Mitchell, E. M., Perkins, N. J., Cole, S. R., Tchetgen Tchetgen, E. J., Sun, B., & Schisterman, E. F. (2018). Multiple imputation for incomplete data in epidemiologic studies. *American Journal of Epidemiology*, *187*(3), 576–584.

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: an International Journal of Obstetrics & Gynaecology*, *125*(13), 1716.

Harner, H. M., & Riley, S. (2013). The impact of incarceration on women's mental health: Responses from women in a maximum-security prison. *Qualitative Health Research*, *23*(1), 26–42.

Harrell, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, *15*(4), 361–387.

Harrer, M., Cuijpers, P., A, F. T., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman & Hall/CRC Press.

Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E., & Finkelstein, J. (2006). The use and interpretation of quasi-experimental studies in medical informatics. *Journal of the American Medical Informatics Association*, *13*(1), 16–23.

Harzke, A. J., Baillargeon, J., Baillargeon, G., Henry, J., Olvera, R. L., Torrealday, O., Penn, J. V., & Parikh, R. (2012). Prevalence of psychiatric disorders in the Texas juvenile correctional system. *Journal of Correctional Health Care*, *18*(2), 143–157.

Hawton, K., Linsell, L., Adeniji, T., Sariaslan, A., & Fazel, S. (2014). Self-harm in prisons in England and Wales: An epidemiological study of prevalence, risk factors, clustering, and subsequent suicide. *Lancet*, *383*(9923), 1147–1154.

Hedrich, D., Alves, P., Farrell, M., Stöver, H., Møller, L., & Mayet, S. (2012). The effectiveness of opioid maintenance treatment in prison settings: A systematic review. *Addiction*, *107*(3), 501–517.

Hein, S., Weeland, J., Square, A., Haeffel, G. J., Chapman, J., Macomber, D., Lee, M., Geib, C. F., & Grigorenko, E. L. (2020). Effectiveness of a social problem solving training in youth in detention or on probation: An RCT and pre-post community implementation. *International Journal of Law and Psychiatry*, *72*, 101626.

Held, U., Kessels, A., Garcia Aymerich, J., Basagaña, X., Ter Riet, G., Moons, K. G., & Puhan, M. A. (2016). Methods for handling missing variables in risk prediction models. *American Journal of Epidemiology*, *184*(7), 545–551.

Henwood, K. S., Chou, S., & Browne, K. D. (2015). A systematic review and meta-analysis on the effectiveness of CBT informed anger management. *Aggression and Violent Behavior*, *25B*, 280–292.

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*.

Higgins, J. P., & Green, S. (2019). *Cochrane handbook for systematic reviews of interventions* (Second edition). John Wiley & Sons.

Hippisley-Cox, J., Coupland, C., & Brindle, P. (2017). Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: Prospective cohort study. *BMJ*, *357*, j2099.

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., May, M., & Brindle, P. (2007). Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study. *BMJ*, *335*, 136.

Hirschtritt, M. E., & Binder, R. L. (2017). Interrupting the mental illness–incarceration-recidivism cycle. *JAMA*, *317*(7), 695–696.

HM Inspectorate of Probation and HM Inspectorate of Prisons. (2016). *An inspection of through the gate resettlement services for short-term prisoners*. Her Majesty's Inspectorate of Probation. Manchester, UK.

Hollander, H. E., & Turner, F. D. (1985). Characteristics of incarcerated delinquents: Relationship between development disorders, environmental and family factors, and patterns of offense and recidivism. *Journal of the American Academy of Child Psychiatry*, *24*(2), 221–226.

Homant, R. J. (1986). Ten years after: A follow-up of therapy effectiveness. *Journal of Offender Counseling Services Rehabilitation*, *10*(3), 51–57.

Hoogland, J., IntHout, J., Belias, M., Rovers, M. M., Riley, R. D., E. Harrell Jr, F., Moons, K. G., Debray, T. P., & Reitsma, J. B. (2021). A tutorial on individualized treatment effect prediction from randomized trials with a binary endpoint. *Statistics in Medicine*, *40*(26), 5961–5981.

Hopkin, G., Evans-Lacko, S., Forrester, A., Shaw, J., & Thornicroft, G. (2018). Interventions at the transition from prison to the community for prisoners with mental illness: A systematic review. *Administration and Policy in Mental Health and Mental Health Services Research*, *45*(4), 623–634.

Hou, X.-H., Feng, L., Zhang, C., Cao, X.-P., Tan, L., & Yu, J.-T. (2019). Models for predicting risk of dementia: A systematic review. *Journal of Neurology, Neurosurgery & Psychiatry*, *90*(4), 373–379.

Howard, P. (2006). *The offender assessment system: An evaluation of the second pilot.* London, UK, Home Office.

Howard, P., & Dixon, L. (2011). Developing an empirical classification of violent offences for use in the prediction of recidivism in England and Wales. *Journal of Aggression, Conflict and Peace Research*, *3*(3), 141–154.

Howard, P. D., & Dixon, L. (2012). The construction and validation of the OASys Violence Predictor: advancing violence risk assessment in the English and Welsh correctional services. *Criminal Justice and Behavior*, *39*(3), 287–307.

Hughes, N., Ungar, M., Fagan, A., Murray, J., Atilola, O., Nichols, K., Garcia, J., & Kinner, S. (2020). Health determinants of adolescent criminalisation. *Lancet Child & Adolescent Health*, *4*(2), 151–162.

Humber, N., Webb, R., Piper, M., Appleby, L., & Shaw, J. (2013). A national case-control study of risk factors among prisoners in England and Wales. *Social Psychiatry and Psychiatric Epidemiology*, *48*(7), 1177–1185.

Igoumenou, A., Kallis, C., & Coid, J. (2015). Treatment of psychosis in prisons and violent recidivism. *BJPsych Open*, *1*(2), 149–157.

Indig, D., Vecchiato, C., Haysom, L., Beilby, R., Carter, J., Champion, U., Gaskin, C., Heller, E., Kumar, S., Mamone, N., et al. (2009). *2009 New South Wales young people in custody health survey: Full report.* Justice Health and Juvenile Justice. Sydney, Australia.

Innes, C. A., & Everett, R. S. (2008). Factors and conditions influencing the use of research by the criminal justice system. *Western Criminology Review*, *9*(1), 49–58.

Jack, H. E., Fricchione, G., Chibanda, D., Thornicroft, G., Machando, D., & Kidia, K. (2018). Mental health of incarcerated people: A global call to action. *Lancet Psychiatry*, *5*(5), 391–392.

Jackson, D., Bowden, J., & Baker, R. (2010). How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? *Journal of Statistical Planning and Inference*, *140*(4), 961–970.

Jalali, M. S., DiGennaro, C., & Sridhar, D. (2020). Transparency assessment of COVID-19 models. *Lancet Global Health*, *8*(12), e1459–e1460.

Janssen, K. J., Donders, A. R. T., Harrell Jr, F. E., Vergouwe, Y., Chen, Q., Grobbee, D. E., & Moons, K. G. (2010). Missing covariate data in medical research: To impute is better than to ignore. *Journal of Clinical Epidemiology*, *63*(7), 721–727.

Jenkins, R., Bhugra, D., Meltzer, H., Singleton, N., Bebbington, P., Brugha, T., Coid, J., Farrell, M., Lewis, G., & Paton, J. (2005). Psychiatric and social aspects of suicidal behaviour in prisons. *Psychological Medicine*, *35*(2), 257–269.

Jennings, W. G., Piquero, A. R., & Reingle, J. M. (2012). On the overlap between victimization and offending: A review of the literature. *Aggression and Violent Behavior*, *17*(1), 16–26.

Jolliffe, D., & Farrington, D. P. (2007). *A systematic review of the national and international evidence on the effectiveness of interventions with violent offenders*. Ministry of Justice London, England. London, UK.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

Justice, A. C., Covinsky, K. E., & Berlin, J. A. (1999). Assessing the generalizability of prognostic information. *Annals of Internal Medicine*, *130*(6), 515–524.

Kaivanto, K. (2008). Maximization of the sum of sensitivity and specificity as a diagnostic cutpoint criterion. *Journal of Clinical Epidemiology*, *61*(5), 517–518.

Kanis, S., Messner, S. F., Eisner, M. P., & Heitmeyer, W. (2017). A cautionary note about the use of estimated homicide data for cross-national research. *Homicide Studies*, *21*(4), 312–324.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481.

Karnik, N. S., Soller, M. V., Redlich, A., Silverman, M. A., Kraemer, H. C., Haapanen, R., & Steiner, H. (2010). Prevalence differences of psychiatric disorders among youth after nine months or more of incarceration by race/ethnicity and age. *Journal of Health Care for the Poor and Underserved*, *21*(1), 237–250.

Kashani, J. H., Manning, G. W., McKnew, D. H., Cytryn, L., Simonds, J. F., & Wooderson, P. C. (1980). Depression among incarcerated delinquents. *Psychiatry Research*, *3*(2), 185–191.

Kelly, U. A. (2011). Theories of intimate partner violence: From blaming the victim to acting against injustice: Intersectionality as an analytic framework. *Advances in Nursing Science*, *34*(3), E29–E51.

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, *52*(12), 1048–1060.

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Cai, T., Ebert, D. D., Hwang, I., Li, J., de Jonge, P., et al. (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular Psychiatry*, *21*(10), 1366–1371.

Kim, J. I., Kim, B., Kim, B.-N., Hong, S.-B., Lee, D. W., Chung, J.-Y., Choi, J. Y., Choi, B.-S., Oh, Y.-R., & Youn, M. (2017). Prevalence of psychiatric disorders, comorbidity patterns, and repeat offending among male juvenile detainees in South Korea: A cross-sectional study. *Child and Adolescent Psychiatry and Mental Health*, *11*(1), 1–9.

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

Kinner, S. A., Forsyth, S., & Williams, G. (2013). Systematic review of record linkage studies of mortality in ex-prisoners: Why (good) methods matter. *Addiction*, *108*(1), 38–49.

Kinner, S. A., & Young, J. T. (2018). Understanding and improving the health of people who experience incarceration: An overview and synthesis. *Epidemiologic Reviews*, *40*(1), 4–11.

Kirkbride, J. B., Fearon, P., Morgan, C., Dazzan, P., Morgan, K., Tarrant, J., Lloyd, T., Holloway, J., Hutchinson, G., Leff, J. P., et al. (2006). Heterogeneity in incidence rates of schizophrenia and other psychotic syndromes: Findings from the 3-center AeSOP study. *Archives of General Psychiatry*, *63*(3), 250–258.

Klonsky, D. E. (2011). Non-suicidal self-injury in United States adults: Prevalence, sociodemographics, topography and functions. *Psychological Medicine*, *41*(9), 1981–1986.

Koehler, J. A., Lösel, F., Akoensi, T. D., & Humphreys, D. K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in europe. *Journal of Experimental Criminology*, *9*(1), 19–43.

Köhler, D., Heinzen, H., Hinrichs, G., & Huchzermeier, C. (2009). The prevalence of mental disorders in a German sample of male incarcerated juvenile offenders. *International Journal of Offender Therapy and Comparative Criminology*, *53*(2), 211–227.

Komalasari, R., Wilson, S., & Haw, S. (2021). A systematic review of qualitative evidence on barriers to and facilitators of the implementation of opioid agonist treatment (OAT) programmes in prisons. *International Journal of Drug Policy*, *87*, 102978.

Komarovskaya, I. A., Booker Loper, A., Warren, J., & Jackson, S. (2011). Exploring gender differences in trauma exposure and the emergence of symptoms of PTSD among incarcerated men and women. *Journal of Forensic Psychiatry & Psychology*, *22*(3), 395–410.

Kouyoumdjian, F. G., McIsaac, K. E., Foran, J. E., & Matheson, F. I. (2017). Canadian Institutes of Health Research funding of prison health research: A descriptive study. *Canadian Medical Association Open Access Journal*, *5*(1), E14–E18.

Kouyoumdjian, F. G., McIsaac, K. E., Liauw, J., Green, S., Karachiwalla, F., Siu, W., Burkholder, K., Binswanger, I., Kiefer, L., Kinner, S. A., et al. (2015). A systematic review of randomized controlled trials of interventions to improve the health of persons during imprisonment and in the year after release. *American Journal of Public Health*, *105*(4), e13–e33.

Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health.* World Health Organization. Geneva, Switzerland.

Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *Lancet*, *360*(9339), 1083–1088.

Kubiak, S., Fedock, G., Kim, W. J., & Bybee, D. (2016). Long-term outcomes of a RCT intervention study for women with violent crimes. *Journal of the Society for Social Work and Research*, *7*(4), 661–679.

Kuitunen, I., Ponkilainen, V. T., Uimonen, M. M., Eskelinen, A., & Reito, A. (2021). Testing the proportional hazards assumption in Cox regression and dealing with possible non-proportionality in total joint arthroplasty research: Methodological perspectives and review. *BMC Musculoskeletal Disorders*, *22*(1), 1–7.

Kuo, E. S., Stoep, A. V., & Stewart, D. G. (2005). Using the short mood and feelings questionnaire to detect depression in detained adolescents. *Assessment*, *12*(4), 374–383.

Labarere, J., Bertrand, R., & Fine, M. J. (2014). How to derive and validate clinical prediction models for use in intensive care medicine. *Intensive Care Medicine*, *40*(4), 513–527.

Lader, D., Singleton, N., & Meltzer, H. (2000). *Psychiatric morbidity among young offenders in England and Wales.* Office for National Statistics. London, UK.

Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive–behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, *1*(4), 451–476.

Långström, N., Enebrink, P., Laurén, E.-M., Lindblom, J., Werkö, S., & Hanson, R. K. (2013). Preventing sexual abusers of children from reoffending: Systematic review of medical and psychological interventions. *BMJ*, *347*, f4630.

Latkin, C. A., Edwards, C., Davey-Rothwell, M. A., & Tobin, K. E. (2017). The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. *Addictive Behaviors*, *73*, 133–136.

Lattimore, P. K., Witte, A. D., & Baker, J. R. (1990). Experimental assessment of the effect of vocational training on youthful property offenders. *Evaluation Review*, *14*(2), 115–133.

Laupacis, A., Sekar, N., & Stiell, l. G. (1997). Clinical prediction rules: A review and suggested modifications of methodological standards. *JAMA*, *277*, 488–494.

Lavis, J. N., Posada, F. B., Haines, A., & Osei, E. (2004). Use of research to inform public policymaking. *Lancet*, *364*(9445), 1615–1621.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lederman, C. S., Dakof, G. A., Larrea, M. A., & Li, H. (2004). Characteristics of adolescent females in juvenile detention. *International Journal of Law and Psychiatry*, *27*(4), 321–337.

Leeman, L. W., Gibbs, J. C., & Fuller, D. (1993). Evaluation of a multi-component group treatment program for juvenile delinquents. *Aggressive Behavior*, *19*(4), 281–292.

Lennox, C., Bell, V., O'Malley, K., Shaw, J., & Dolan, M. (2013). A prospective cohort study of the changing mental health needs of adolescents in custody. *BMJ Open*, *3*(3), e002358.

Levey, D. F., Niculescu, E. M., Le-Niculescu, H., Dainton, H., Phalen, P. L., Ladd, T. B., Weber, H., Belanger, E., Graham, D. L., Khan, F. N., et al. (2016). Towards understanding and predicting suicidality in women: Biomarkers and clinical risk assessment. *Molecular Psychiatry*, *21*(6), 768–785.

Levine, A. C., Glavis-Bloom, J., Modi, P., Nasrin, S., Atika, B., Rege, S., Robertson, S., Schmid, C. H., & Alam, N. H. (2016). External validation of the DHAKA score and comparison with the current IMCI algorithm for the assessment of dehydration in children with diarrhoea: A prospective cohort study. *Lancet Global Health*, *4*(10), e744–e751.

Lewis, R. V. (1983). Scared straight—California style: evaluation of the San Quentin SQUIRES program. *Criminal Justice and Behavior*, *10*(2), 209–226.

Lilienfeld, D. E., Stolley, P. D., & Lilienfeld, A. M. (1994). *Foundations of epidemiology.* Oxford University Press.

Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics*, *74*(3), 785–794.

Lindblad, F., Isaksson, J., Heiskala, V., Koposov, R., & Ruchkin, V. (2015). Comorbidity and behavior characteristics of Russian male juvenile delinquents with ADHD and conduct disorder. *Journal of Attention Disorders*, *24*(7), 1070–1077.

Linden, R., & Perry, L. (1984). An evaluation of a prison education program. *Canadian Journal of Criminology*, *26*(1), 65–73.

Lindforss, L., & Magnusson, D. (1997). Solution-focused therapy in prison. *Contemporary Family Therapy*, *19*(1), 89–103.

Link, N. W., Ward, J. T., & Stansfield, R. (2019). Consequences of mental and physical health for reentry and recidivism: Toward a health-based model of desistance. *Criminology*, *57*(3), 544–573.

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.

Lipsey, M. W., Landenberger, N. A., & Wilson, S. J. (2007). Effects of cognitive-behavioral programs for criminal offenders. *Campbell Systematic Reviews*, *3*(1), 1–27.

Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful intervention* (pp. 313–345). Sage Publications.

Little, R. J. A., & Rubin, D. B. (2020). *Statistical analysis with missing data* (3rd ed). Wiley.

Livanou, M., Furtado, V., Winsper, C., Silvester, A., & Singh, S. P. (2019). Prevalence of mental disorders and symptoms among incarcerated youth: A meta-analysis of 30 studies. *International Journal of Forensic Mental Health*, *18*(4), 400–414.

Lobmaier, P. P., Kunøe, N., & Waal, H. (2010). Treatment research in prison: Problems and solutions in a randomized trial. *Addiction Research & Theory*, *18*(1), 1–13.

Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology*, *1*(1), 117–146.

Lucas, P. J., & Staines, J. (2022). *Supporting the youngest children in the youth justice system: What works to reduce offending and improve outcomes?* Local Government Association. London, UK.

Luxton, D. D., Skopp, N. A., & Maguen, S. (2010). Gender differences in depression and PTSD symptoms following combat exposure. *Depression and Anxiety*, *27*(11), 1027–1033.

Macdonald, G. (2002). Violence and health: the ultimate public health challenge. *Health Promotion International*, *17*(4), 293–295.

Mallett, S., Halligan, S., Thompson, M., Collins, G. S., & Altman, D. G. (2012). Interpreting diagnostic accuracy studies for patient care. *BMJ*, *345*.

Malouf, E. T., Youman, K., Stuewig, J., Witt, E. A., & Tangney, J. P. (2017). A pilot RCT of a values-based mindfulness group intervention with jail inmates: Evidence for reduction in post-release risk behavior. *Mindfulness*, *8*(3), 603–614.

Maret-Ouda, J., Tao, W., Wahlin, K., & Lagergren, J. (2017). Nordic registry-based cohort studies: Possibilities and pitfalls when combining Nordic registry data. *Scandinavian Journal of Public Health*, *45*(17_suppl), 14–19.

Marshall, A., Altman, D. G., Holder, R. L., & Royston, P. (2009). Combining estimates of interest in prognostic modelling studies after multiple imputation: Current practice and guidelines. *BMC Medical Research Methodology*, *9*, 57.

Mascayano, F., Armijo, J. E., & Yang, L. H. (2015). Addressing stigma relating to mental illness in low-and middle-income countries. *Frontiers in Psychiatry*, *11*(6), 38.

Mata, D. A., Ramos, M. A., Bansal, N., Khan, R., Guille, C., Di Angelantonio, E., & Sen, S. (2015). Prevalence of depression and depressive symptoms among resident physicians: A systematic review and meta-analysis. *JAMA*, *314*(22), 2373–2383.

Mavridis, D., & Salanti, G. (2014). How to assess publication bias: Funnel plot, trim-and-fill method and selection models. *Evidence-Based Mental Health*, *17*, 30.

McLernon, D. J., Giardiello, D., Van Calster, B., Wynants, L., van Geloven, N., van Smeden, M., Therneau, T., Steyerberg, E. W., & topic groups 6 and 8 of the STRATOS Initiative. (2022). Assessing performance and clinical usefulness in prediction models with survival outcomes: Practical guidance for Cox proportional hazards models. *medRxiv*.

McReynolds, L. S., Schwalbe, C. S., & Wasserman, G. A. (2010). The contribution of psychiatric disorder to juvenile recidivism. *Criminal Justice and Behavior*, *37*(2), 204–216.

Meehan, A. J., Lewis, S. J., Fazel, S., Fusar-Poli, P., Steyerberg, E. W., Stahl, D., & Danese, A. (2022). Clinical prediction models in psychiatry: A systematic review of two decades of progress and challenges. *Molecular Psychiatry*, *27*, 2700–2708.

Mercy, J. A., Hillis, S. D., Butchart, A., Bellis, M. A., Ward, C. L., Fang, X., & Rosenberg, M. L. (2017). Interpersonal violence: Global impact and paths to prevention. In C. N. Mock, R. Nugent, O. Kobusingye, & K. R. S. Smith (Eds.), *Injury prevention and environmental health. 3rd edition.* (pp. 71–96). The International Bank for Reconstruction; Development/The World Bank.

Merikangas, K. R., He, J.-P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., Benjet, C., Georgiades, K., & Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(10), 980–989.

Merten, E. C., Cwik, J. C., Margraf, J., & Schneider, S. (2017). Overdiagnosis of mental disorders in children and adolescents (in developed countries). *Child and Adolescent Psychiatry and Mental Health*, *11*(5).

Messina, N., Grella, C. E., Cartier, J., & Torres, S. (2010). A randomized experimental study of gender-responsive substance abuse treatment for women in prison. *Journal of Substance Abuse Treatment*, *38*(2), 97–107.

Mitchell, P., & Shaw, J. (2011). Factors affecting the recognition of mental health problems among adolescent offenders in custody. *Journal of Forensic Psychiatry & Psychology*, *22*(3), 381–394.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*(4), 674–701.

Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin longitudinal study*. Cambridge University Press.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269.

Molero, Y., Zetterqvist, J., Binswanger, I. A., Hellner, C., Larsson, H., & Fazel, S. (2018). Medications for alcohol and opioid use disorders and risk of suicidal behavior, accidental overdoses, and crime. *American Journal of Psychiatry*, *175*(10), 970–978.

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology*, *12*, 489–513.

Moons, K. G., Kengne, A. P., Grobbee, D. E., Royston, P., Vergouwe, Y., Altman, D. G., & Woodward, M. (2012). Risk prediction models: II. external validation, model updating, and impact assessment. *Heart*, *98*(9), 691–698.

Moore, K. E., Roberts, W., Reid, H. H., Smith, K. M., Oberleitner, L. M., & McKee, S. A. (2019). Effectiveness of medication assisted treatment for opioid use in prison and jail settings: A meta-analysis and systematic review. *Journal of Substance Abuse Treatment*, *99*, 32–43.

Mudumbai, S. C., & Rashidi, P. (2021). Linking preoperative and intraoperative data for risk prediction: More answers or just more data? *JAMA Network Open*, *4*(3), e212547.

Mundt, A. P., Baranyi, G., Gabrysch, C., & Fazel, S. (2018). Substance use during imprisonment in low- and middle-income countries. *Epidemiologic Reviews*, *40*(1), 70–81.

Murray, J., & Atilola, O. (2020). Determinants of youth crime in low-income and middle-income countries. *Lancet Child & Adolescent Health*, *4*(2), 96–98.

Murray, J., Farrington, D. P., & Sekol, I. (2012). Children's antisocial behavior, mental health, drug use, and educational performance after parental incarceration: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 175–210.

National Research Council. (2008). *Parole, desistance from crime, and community integration*. The National Academies Press.

Navarro, C. L. A., Damen, J. A., Takada, T., Nijman, S. W., Dhiman, P., Ma, J., Collins, G. S., Bajpai, R., Riley, R. D., Moons, K. G., et al. (2021). Risk of bias in studies on prediction models developed using supervised machine learning techniques: Systematic review. *BMJ*, *375*, n2281.

Neufeld, S. A. (2022). The burden of young people's mental health conditions in Europe: No cause for complacency. *Lancet Regional Health–Europe*, *16*, 100364.

Newton, A., May, X., Eames, S., & Ahmad, M. (2019). *Economic and social costs of reoffending*. Ministry of Justice. London, UK.

Nguyen, T.-L., Collins, G. S., Landais, P., & Le Manach, Y. (2020). Counterfactual clinical prediction models could help to infer individualized treatment effects in randomized controlled trials—an illustration with the international stroke trial. *Journal of Clinical Epidemiology*, *125*, 47–56.

Nicol, R., Stretch, D., Whitney, I., Jones, K., Garfield, P., Turner, K., & Stanion, B. (2000). Mental health needs and services for severely troubled and troubling young people including young offenders in an NHS region. *Journal of Adolescence*, *23*(3), 243–261.

Nielsen, A. L., Scarpitti, F. R., & Inciardi, J. A. (1996). Integrating the therapeutic community and work release for drug-involved offenders: The CREST program. *Journal of Substance Abuse Treatment*, *13*(4), 349–358.

Nikolakopoulou, A., Mavridis, D., & Salanti, G. (2014). Demystifying fixed and random effects meta-analysis. *Evidence-Based Mental Health*, *17*(2), 53–57.

Niolon, P. H., Treves-Kagan, S., Dahlberg, L. L., & Mercy, J. A. (2020). The evolution of interpersonal violence research and prevention across the lifespan in the United States: The past, present, and future. In R. Geffner, J. W. White, L. K. Hamberger, A. Rosenbaum, V. Vaughan-Eden, & V. I. Vieth (Eds.), *Handbook of interpersonal violence and abuse across the lifespan* (pp. 1–28). Springer.

Noble, S., McLennan, D., Noble, M., Plunkett, E., Gutacker, N., Silk, M., & Wright, G. (2019). *The English Indices of Deprivation 2019 — research report.* Ministry of Housing, Communities & Local Government. London, UK.

Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2006). Prevalence, subtypes, and correlates of DSM-IV conduct disorder in the National Comorbidity Survey Replication. *Psychological Medicine*, *36*(5), 699–710.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.

Nygaard Andersen, S., & Skardhamar, T. (2017). Pick a number: Mapping recidivism measures and their consequences. *Crime & Delinquency*, *63*(5), 613–635.

Nygaard Andersen, S., & Telle, K. (2022). Better out than in? The effect on recidivism of replacing incarceration with electronic monitoring in Norway. *European Journal of Criminology*, *19*(1), 55–76.

Ochodo, E. A., de Haan, M. C., Reitsma, J. B., Hooft, L., Bossuyt, P. M., & Leeflang, M. M. (2013). Overinterpretation and misreporting of diagnostic accuracy studies: Evidence of "spin". *Radiology*, *267*(2), 581–588.

Office for National Statistics (UK). (2021). *The nature of violent crime in England and Wales: year ending March 2020.* Retrieved April 29, 2022, from https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/thenatureofviolentcrimeinenglandandwales/yearendingmarch2020

O'Malley, P. M., Bachman, J. G., & Johnston, L. D. (1983). Reliability and consistency in self-reports of drug use. *International Journal of the Addictions*, *18*(6), 805–824.

Ortmann, R. (2000). The effectiveness of social therapy in prison—a randomized experiment. *Crime & Delinquency*, *46*(2), 214–232.

Pache, S. (2020). A history of interpersonal violence: Raising public concern. In R. Geffner, J. W. White, L. K. Hamberger, A. Rosenbaum, V. Vaughan-Eden, & V. I. Vieth (Eds.), *Handbook of interpersonal violence and abuse across the lifespan* (pp. 1–22). Springer.

Paoli, B., Haggard, L., & Shah, G. (2002). *Confidence intervals in public health.* Office of Public Health Assessment, Utah Department of Health. Utah, US.

Papalia, N., Spivak, B., Daffern, M., & Ogloff, J. R. P. (2020). Are psychological treatments for adults with histories of violent offending associated with change in dynamic risk factors? A meta-analysis of intermediate treatment outcomes. *Criminal Justice and Behavior*, *47*(12), 1585–1608.

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

Parikh, N. I., Pencina, M. J., Wang, T. J., Benjamin, E. J., Lanier, K. J., Levy, D., D'Agostino Sr, R. B., Kannel, W. B., & Vasan, R. S. (2008). A risk score for predicting near-term incidence of hypertension: The Framingham Heart Study. *Annals of Internal Medicine*, *148*(2), 102–110.

Parker, G., & Brotchie, H. (2010). Gender differences in depression. *International Review of Psychiatry*, *22*(5), 429–436.

Patton, G. C., Sawyer, S. M., Santelli, J. S., Ross, D. A., Afifi, R., Allen, N. B., Arora, M., Azzopardi, P., Baldwin, W., Bonell, C., et al. (2016). Our future: A Lancet commission on adolescent health and wellbeing. *Lancet*, *387*(10036), 2423–2478.

Pavlou, M., Qu, C., Omar, R. Z., Seaman, S. R., Steyerberg, E. W., White, I. R., & Ambler, G. (2021). Estimation of required sample size for external validation of risk models for binary outcomes. *Statistical Methods in Medical Research*, *30*(10), 2187–2206.

Pearson, F. S., Lipton, D. S., Cleland, C. M., & Yee, D. S. (2002). The effects of behavioral/cognitive-behavioral programs on recidivism. *Crime & Delinquency*, *48*(3), 476–496.

Pencina, M. J., Goldstein, B. A., & D'Agostino, R. B. (2020). Prediction models—development, evaluation, and clinical application. *New England Journal of Medicine*, *382*(17), 1583–1586.

Penn, J. V., Esposito, C. L., Schaeffer, L. E., Fritz, G. K., & Spirito, A. (2003). Suicide attempts and self-mutilative behavior in a juvenile correctional facility. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(7), 762–769.

Pérez, D. M., Gover, A. R., Tennyson, K. M., & Santos, S. D. (2010). Individual and institutional characteristics related to inmate victimization. *International Journal of Offender Therapy and Comparative Criminology*, *54*(3), 378–394.

Perry, A. E., Martyn-St James, M., Burns, L., Hewitt, C., Glanville, J. M., Aboaja, A., Thakkar, P., Kumar, K. M. S., Pearson, C., Wright, K., et al. (2019a). Interventions for drug-using offenders with co-occurring mental health problems. *Cochrane Database of Systematic Reviews*, (10).

Perry, A. E., Martyn-St James, M., Burns, L., Hewitt, C., Glanville, J. M., Aboaja, A., Thakkar, P., Kumar, K. M. S., Pearson, C., & Wright, K. (2019b). Interventions for female drug-using offenders. *Cochrane Database of Systematic Reviews*, (12).

Perry, A. E., Woodhouse, R., Neilson, M., Martyn St James, M., Glanville, J., Hewitt, C., & Trépel, D. (2016). Are non-pharmacological interventions effective in reducing drug use and criminality? A systematic and meta-analytical review with an economic appraisal of these interventions. *International Journal of Environmental Research and Public Health*, *13*(10), 966.

Persons, R. W. (1967). Relationship between psychotherapy with institutionalized boys and subsequent community adjustment. *Journal of Consulting Psychology*, *31*(2), 137–141.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *JAMA*, *295*(6), 676–680.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine*, *26*(25), 4544–4562.

Petersilia, J. (2011). Beyond the prison bubble. *The Wilson Quarterly (1976-)*, *35*(1), 50–55.

Pieper, D., & Rombey, T. (2022). Where to prospectively register a systematic review. *Systematic Reviews*, *11*(1), 1–8.

Plattner, B., Steiner, H., The, S. S., Kraemer, H. C., Bauer, S. M., Kindler, J., Friedrich, M. H., Kasper, S., & Feucht, M. (2009). Sex-specific predictors of criminal recidivism in a representative sample of incarcerated youth. *Comprehensive Psychiatry*, *50*(5), 400–407.

Pliszka, S. R., Sherman, J. O., Barrow, M. V., & Irick, S. (2000). Affective disorder in juvenile offenders: A preliminary study. *American Journal of Psychiatry*, *157*(1), 130–132.

Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C., & Rohde, L. A. (2014). ADHD prevalence estimates across three decades: An updated systematic review and meta-regression analysis. *International Journal of Epidemiology*, *43*(2), 434–442.

Poole, C., & Greenland, S. (1999). Random-effects meta-analyses are not always conservative. *American Journal of Epidemiology*, *150*(5), 469–475.

Pratt, D., Appleby, L., Piper, M., Webb, R., & Shaw, J. (2010). Suicide in recently released prisoners: A case-control study. *Psychological Medicine*, *40*(5), 827–835.

Prendergast, M. L., Hall, E. A., Wexler, H. K., Melnick, G., & Cao, Y. (2004). Amity prison-based therapeutic community: 5-year outcomes. *The Prison Journal*, *84*(1), 36–60.

Proctor, S. L., Hoffmann, N. G., & Allison, S. (2012). The effectiveness of interactive journaling in reducing recidivism among substance-dependent jail inmates. *International Journal of Offender Therapy and Comparative Criminology*, *56*(2), 317–332.

Quina, K., Garis, A. V., Stevenson, J., Garrido, M., Brown, J., Richman, R., Renzi, J., Fox, J., & Mitchell, K. (2007). Through the bullet-proof glass: Conducting research in prison settings. *Journal of Trauma & Dissociation*, *8*(2), 123–139.

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Ramesh, T., Igoumenou, A., Montes, M. V., & Fazel, S. (2018). Use of risk assessment instruments to predict violence in forensic psychiatric hospitals: A systematic review and meta-analysis. *European Psychiatry*, *52*, 47–53.

Ramspek, C. L., Jager, K. J., Dekker, F. W., Zoccali, C., & van Diepen, M. (2021). External validation of prognostic models: What, why, how, when and where? *Clinical Kidney Journal*, *14*(1), 49–58.

Raphael, S. (2011). Incarceration and prisoner reentry in the United States. *The Annals of the American Academy of Political and Social Science*, *635*(1), 192–215.

Rezansoff, S. N., Moniruzzaman, A., Gress, C., & Somers, J. M. (2013). Psychiatric diagnoses and multiyear criminal recidivism in a Canadian provincial offender population. *Psychology, Public Policy, and Law*, *19*(4), 443–453.

Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: The violence risk appraisal guide—revised (VRAG-R). *Psychological Assessment*, *25*(3), 951.

Robertson, A., & Husain, J. (2001). *Prevalence of mental illness and substance abuse disorders among incarcerated juvenile offenders.* Mississippi Department of Public Safety. Mississipi, US.

Robinson, D. (1995). *The impact of cognitive skills training on post-release recidivism among Canadian federal offenders.* Correctional Service Canada, Correctional Research & Development. Ottawa, Ontario.

Rodriguez, F. S., & Roehr, S. (2020). Challenges in dementia risk prediction in low-income and middle-income countries. *Lancet Global Health*, *8*(4), e458–e459.

Rosenberg, M. L. (1985). *Violence as a public health problem: Background papers for the surgeon general's workshop on violence and public health.* US Department of Health, Human Services, and US Department of Justice. Leesburg, VA.

Royston, P., & Altman, D. G. (2013). External validation of a Cox prognostic model: Principles and methods. *BMC Medical Research Methodology*, *13*(1), 1–15.

RStudio Team. (2021). *Rstudio: Integrated development environment for R*. RStudio, Inc. Boston, MA. http://www.rstudio.com/

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Ruchkin, V. V., Schwab-Stone, M., Koposov, R., Vermeiren, R., & Steiner, H. (2002). Violence exposure, posttraumatic stress, and personality in juvenile delinquents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(3), 322–329.

Russolillo, A., Moniruzzaman, A., McCandless, L. C., Patterson, M., & Somers, J. M. (2018). Associations between methadone maintenance treatment and crime: A 17-year longitudinal cohort study of Canadian provincial offenders. *Addiction*, *113*(4), 656–667.

Sacks, J. Y., McKendrick, K., & Hamilton, Z. (2012). A randomized clinical trial of a therapeutic community treatment for female inmates: Outcomes at 6 and 12 months after prison release. *Journal of Addictive Diseases*, *31*(3), 258–269.

Sacks, S., Sacks, J. Y., McKendrick, K., Banks, S., & Stommel, J. (2004). Modified TC for MICA offenders: Crime outcomes. *Behavioral Sciences & the Law*, *22*(4), 477–501.

Salazar de Pablo, G., Studerus, E., Vaquerizo-Serrano, J., Irving, J., Catalan, A., Oliver, D., Baldwin, H., Danese, A., Fazel, S., Steyerberg, E. W., et al. (2021). Implementing precision psychiatry: A systematic review of individualized prediction models for clinical practice. *Schizophrenia Bulletin*, *47*(2), 284–297.

Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.

Sanci, L. (2019). Understanding and responding to the long-term burdens of childhood sexual abuse. *Lancet Psychiatry*, *6*(10), 795–797.

Sarabipour, S., Debat, H. J., Emmott, E., Burgess, S. J., Schwessinger, B., & Hensel, Z. (2019). On the value of preprints: An early career researcher perspective. *PLOS Biology*, *17*(2), e3000151.

Sariaslan, A., Arseneault, L., Larsson, H., Lichtenstein, P., & Fazel, S. (2020). Risk of subjection to violence and perpetration of violence in persons with psychiatric disorders in Sweden. *JAMA Psychiatry*, *77*(4), 359–367.

Sawyer, S. M., Afifi, R. A., Bearinger, L. H., Blakemore, S.-J., Dick, B., Ezeh, A. C., & Patton, G. C. (2012). Adolescence: A foundation for future health. *Lancet*, *379*(9826), 1630–1640.

Sawyer, W. (2019). *Youth confinement: The whole pie 2019*. Retrieved May 21, 2022, from https://www.prisonpolicy.org/reports/youth2019.html

Sawyer, W., & Wagner, P. (2022). *Mass incarceration: The whole pie 2022*. Retrieved May 2, 2022, from https://www.prisonpolicy.org/reports/pie2022.html

Scheyett, A. (2022). Impact of prison security level on mortality. *Lancet Public Health*, *7*(7), e574–e575.

Schmitt, J., & Warner, K. (2010). *Ex-offenders and the labor market*. Center for Economic and Policy Research. Washington, DC.

Schmucker, M., & Lösel, F. (2015). The effects of sexual offender treatment on recidivism: An international meta-analysis of sound quality evaluations. *Journal of Experimental Criminology*, *11*(4), 597–630.

Schnittker, J., Massoglia, M., & Uggen, C. (2012). Out and down: Incarceration and psychiatric disorders. *Journal of Health and Social Behavior*, *53*(4), 448–464.

Schorr, M. T., Reichelt, R. R., Alves, L. P. d. C., Telles, B. d. B., Strapazzon, L., & Telles, L. E. d. B. (2019). Youth homicide: A study of homicide predictor factors in adolescent offenders in custody in the south of Brazil. *Trends in Psychiatry and Psychotherapy*, *41*(3), 292–296.

Schroeder, R. D., Hill, T. D., Haynes, S. H., & Bradley, C. (2011). Physical health and crime among low-income urban women: An application of general strain theory. *Journal of Criminal Justice*, *39*(1), 21–29.

Schubert, C. A., Mulvey, E. P., & Glasheen, C. (2011). Influence of mental health and substance use problems and criminogenic risk on outcomes in serious juvenile offenders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(9), 925–937.

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2017). *Meta-analysis in R*. Springer.

SCORE2 working group and ESC Cardiovascular risk collaboration. (2021). SCORE2 risk prediction algorithms: New models to estimate 10-year risk of cardiovascular disease in Europe. *European Heart Journal*, *42*(25), 2439–2454.

Senior, M., Burghart, M., Yu, R., Kormilitzin, A., Liu, Q., Vaci, N., Nevado-Holgado, A., Pandit, S., Zlodre, J., & Fazel, S. (2020). Identifying predictors of suicide in severe mental illness: A feasibility study of a clinical prediction rule (Oxford Mental Illness and Suicide Tool or OxMIS). *Frontiers in Psychiatry*, *11*, 268.

Senior, M., Fanshawe, T., Fazel, M., & Fazel, S. (2021). Prediction models for child and adolescent mental health: A systematic review of methodology and reporting in recent research. *JCPP Advances*, *1*(3), e12034.

Sestelo, M. (2017). *A short course on survival analysis applied to the financial industry.* Retrieved July 28, 2022, from https://bookdown.org/sestelo/sa_financial/

Shapland, J., Atkinson, A., Atkinson, H., Dignan, J., Edwards, L., Hibbert, J., Howes, M., Johnstone, J., Robinson, G., & Sorsby, A. (2008). *Does restorative justice affect reconviction. The fourth report from the evaluation of three schemes.* Ministry of Justice. London, UK.

Shelton, D. (1998). *Estimates of emotional disorder in detained and committed youth in the Maryland juvenile justice system.* University of Maryland, School of Nursing. Baltimore, ML.

Shi, L., & Lin, L. (2019). The trim-and-fill method for publication bias: Practical guidelines and recommendations based on a large database of meta-analyses. *Medicine*, *98*(23), e15987.

Shillan, D., Sterne, J. A., Champneys, A., & Gibbison, B. (2019). Use of machine learning to analyse routinely collected intensive care unit data: A systematic review. *Critical Care*, *23*(1), 1–11.

Shivrattan, J. L. (1988). Social interactional training and incarcerated juvenile delinquents. *Canadian Journal of Criminology*, *30*(2), 145–163.

Shung, D., Simonov, M., Gentry, M., Au, B., & Laine, L. (2019). Machine learning to predict outcomes in patients with acute gastrointestinal bleeding: A systematic review. *Digestive Diseases and Sciences*, *64*(8), 2078–2087.

Signorini, G., Singh, S. P., Boricevic-Marsanic, V., Dieleman, G., Dodig-Ćurković, K., Franic, T., Gerritsen, S. E., Griffin, J., Maras, A., McNicholas, F., et al. (2017). Architecture and functioning of child and adolescent mental health services: A 28-country survey in Europe. *Lancet Psychiatry*, *4*(9), 715–724.

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., et al. (2014).

International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health*, *13*(3), 193–206.

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, *31*(3), 499–513.

Singh, J. P., Grann, M., & Fazel, S. (2013). Authorship bias in violence risk assessment? a systematic review and meta-analysis. *PLOS One*, *8*(9), e72484.

Singh, J. P., Serper, M., Reinharth, J., & Fazel, S. (2011). Structured assessment of violence risk in schizophrenia and other psychiatric disorders: A systematic review of the validity, reliability, and item content of 10 available instruments. *Schizophrenia Bulletin*, *37*(5), 899–912.

Siontis, G. C., Tzoulaki, I., Castaldi, P. J., & Ioannidis, J. P. (2015). External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *Journal of Clinical Epidemiology*, *68*(1), 25–34.

Sleet, D. A., Baldwin, G., Marr, A., Spivak, H., Patterson, S., Morrison, C., Holmes, W., Peeples, A. B., & Degutis, L. C. (2012). History of injury and violence as public health problems and emergence of the National Center for Injury Prevention and Control at CDC. *Journal of Safety Research*, *43*(4), 233–247.

Snell, K. I., Ensor, J., Debray, T. P., Moons, K. G., & Riley, R. D. (2018). Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C-statistic and calibration measures? *Statistical Methods in Medical Research*, *27*(11), 3505–3522.

Somers, J. M., Moniruzzaman, A., Rezansoff, S. N., Brink, J., & Russolillo, A. (2016). The prevalence and geographic distribution of complex co-occurring disorders: A population study. *Epidemiology and Psychiatric Sciences*, *25*(3), 267–277.

Song, X., Liu, X., Liu, F., & Wang, C. (2021). Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. *International Journal of Medical Informatics*, *151*, 104484.

Sørland, T. O., & Kjelsberg, E. (2009). Mental health among teenage boys remanded to prisoner [mental helse hos varetektsfengslede tenåringsgutter]. *Tidsskr Nor Legeforen [Journal of the Norwegian Medical Association]*, *129*(23), 2472–2475.

Stang, A. (2010). Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *European Journal of Epidemiology*, *25*(9), 603–605.

StataCorp. (2017). *Stata statistical software: Release 17*. StataCorp LP. College Station, TX.

Stein, M. B., Walker, J. R., Hazen, A. L., & Forde, D. R. (1997). Full and partial posttraumatic stress disorder: Findings from a community survey. *The American Journal of Psychiatry*, *154*(8), 1114–1119.

Steinfeld, N. I., Powers, M., & Saltmarsh, K. (2018). *Illinois: The high cost of recidivism*. Illinois Sentencing Policy Advisory Council. Chicago, IL.

Stephan, B. C., Pakpahan, E., Siervo, M., Licher, S., Muniz-Terrera, G., Mohan, D., Acosta, D., Pichardo, G. R., Sosa, A. L., Acosta, I., et al. (2020). Prediction of dementia risk in low-income and middle-income countries (the 10/66 study): An independent external validation of existing models. *Lancet Global Health*, *8*(4), e524–e535.

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., et al. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, *366*, l4898.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046–1055.

Steyerberg, E. W. (2009). *Clinical prediction models: A practical approach to development, validation, and updating*. Springer.

Steyerberg, E. W. (2018). Validation in prediction research: The waste by data splitting. *Journal of Clinical Epidemiology*, *103*, 131–133.

Steyerberg, E. W., Moons, K. G. M., van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., Riley, R. D., Hemingway, H., Altman, D. G., & for the PROGRESS Group. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLOS Medicine*, *10*, 1–9.

Steyerberg, E. W., & Vergouwe, Y. (2014). Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *European Heart Journal*, *35*(29), 1925–1931.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology*, *21*(1), 128–138.

Stogner, J., Gibson, C. L., & Miller, J. M. (2014). Examining the reciprocal nature of the health-violence relationship: Results from a nationally representative sample. *Justice Quarterly*, *31*(3), 473–499.

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., Becker, B. J., Sipe, T. A., Thacker, S. B., et al. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA*, *283*(15), 2008–2012.

Su, T.-L., Jaki, T., Hickey, G. L., Buchan, I., & Sperrin, M. (2018). A review of statistical updating methods for clinical prediction models. *Statistical Methods in Medical Research*, *27*(1), 185–197.

Takahashi, Y., & Evans, L. T. (2018). An application of machine learning for predicting rearrests: Significant predictors for juveniles. *Race and Social Problems*, *10*(1), 42–52.

Teplin, L. A., Abram, K. M., McClelland, G. M., Dulcan, M. K., & Mericle, A. A. (2002). Psychiatric disorders in youth in juvenile detention. *Archives of General Psychiatry*, *59*(12), 1133–1143.

Teplin, L. A., Welty, L. J., Abram, K. M., Dulcan, M. K., & Washburn, J. J. (2012). Prevalence and persistence of psychiatric disorders in youth after detention: A prospective longitudinal study. *Archives of General Psychiatry*, *69*(10), 1031–1043.

Thomas, R., Sanders, S., Doust, J., Beller, E., & Glasziou, P. (2015). Prevalence of attention-deficit/hyperactivity disorder: A systematic review and meta-analysis. *Pediatrics*, *135*(4), e994–e1001.

Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing and combining risks for sexual and violent recidivism. *Annals of the New York Academy of Sciences*, *989*(1), 225–235.

Thurtle, D. R., Greenberg, D. C., Lee, L. S., Huang, H. H., Pharoah, P. D., & Gnanapragasam, V. J. (2019). Individual prognosis at diagnosis in nonmetastatic prostate cancer: Development and external validation of the predict prostate multivariable model. *PLOS Medicine*, *16*(3), e1002758.

Tirupakuzhi Vijayaraghavan, B. K., Priyadarshini, D., Rashan, A., Beane, A., Venkataraman, R., Ramakrishnan, N., Haniffa, R., & Indian Registry of IntenSive care (IRIS) collaborators. (2020). Validation of a simplified risk prediction model using a cloud based critical care registry in a lower-middle income country. *PLOS One*, *15*(12), e0244989.

Tolan, P., & Guerra, N. (1994). *What works in reducing adolescent violence: An empirical review of the field*. The Center for the Study and Prevention of Violence. Boulder, CO.

Tran, N. T., Baggio, S., Dawson, A., O'Moore, É., Williams, B., Bedell, P., Simon, O., Scholten, W., Getaz, L., & Wolff, H. (2018). Words matter: A call for humanizing and respectful language to describe people who experience incarceration. *BMC International Health and Human Rights*, *18*(1), 1–6.

Tripodi, S. J., Bledsoe, S. E., Kim, J. S., & Bender, K. (2011). Effects of correctional-based programs for female inmates: A systematic review. *Research on Social Work Practice*, *21*(1), 15–31.

Ukah, U. V., Payne, B., Lee, T., Magee, L. A., & von Dadelszen, P. (2017). External validation of the fullPIERS model for predicting adverse maternal outcomes in pregnancy hypertension in low-and middle-income countries. *Hypertension*, *69*(4), 705–711.

Ulzen, T. P., Psych, D. C., & Hamilton, H. (1998). The nature and characteristics of psychiatric comorbidity in incarcerated adolescents. *The Canadian Journal of Psychiatry*, *43*(1), 57–63.

UNESCO. (1985). *World congress of youth — final report*. Barcelona, Spain.

US Department of Health, E., & Welfare. (1979). *Healthy people: The Surgeon General's report on health promotion and disease prevention*. US Department of Health, Education, Welfare, Public Health Service, Office of the Assistant Secretary for Health, and Surgeon General. Washington, DC.

Usher, A. M., & Stewart, L. A. (2014). Effectiveness of correctional programs with ethnically diverse offenders: A meta-analytic study. *International Journal of Offender Therapy and Comparative Criminology*, *58*(2), 209–230.

Usher, A. M., Stewart, L. A., & Wilton, G. (2013). Attention deficit hyperactivity disorder in a Canadian prison population. *International Journal of Law and Psychiatry*, *36*(3-4), 311–315.

Van Calster, B., McLernon, D. J., Van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine*, *17*(1), 1–7.

Van Ginneken, E. F. J. C. (2019). The use of risk assessment in sentencing. In J. W. de Keijser, J. V. Roberts, & J. Ryberg (Eds.), *Predictive sentencing: Normative and empirical perspectives* (1st ed., pp. 9–32). Hart Publishing.

van Smeden, M., Reitsma, J. B., Riley, R. D., Collins, G. S., & Moons, K. G. (2021). Clinical prediction models: Diagnosis versus prognosis. *Journal of Clinical Epidemiology*, *132*, 142–145.

van Buuren, S. (2012). *Flexible imputation of missing data*. CRC Press.

Vergouwe, Y., Moons, K. G. M., & Steyerberg, E. W. (2010). External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *American Journal of Epidemiology*, *172*(8), 971–980.

Vergouwe, Y., Steyerberg, E. W., Eijkemans, M. J., & Habbema, J. D. F. (2005). Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *Journal of Clinical Epidemiology*, *58*(5), 475–483.

Visser, S. N., Danielson, M. L., Bitsko, R. H., Holbrook, J. R., Kogan, M. D., Ghandour, R. M., Perou, R., & Blumberg, S. J. (2014). Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003–2011. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(1), 34–46.

von Hippel, P. T. (2015). The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Medical Research Methodology*, *15*(1), 1–8.

Vos, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abolhassani, H., Aboyans, V., Abrams, E. M., Abreu, L. G., Abrigo, M. R. M., Abu-Raddad, L. J., Abushouk, A. I., … Murray, C. J. L. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet*, *396*(10258), 1204–1222.

Vreugdenhil, C., Doreleijers, T. A., Vermeiren, R., Wouters, L. F., & Van Den Brink, W. (2004). Psychiatric disorders in a representative sample of incarcerated boys in the Netherlands. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(1), 97–104.

Waite D., N. J. L. (2002). *Profiles of incarcerated adolescents in Virginia's juvenile correctional centers: Fiscal years 1999-2003.* Virginia Department of Juvenile Justice. Richmond, VA.

Wakai, S., Shelton, D., Trestman, R. L., & Kesten, K. (2009). Conducting research in corrections: Challenges and solutions. *Behavioral Sciences & the Law*, *27*(5), 743–752.

Wang, W., Kiik, M., Peek, N., Curcin, V., Marshall, I. J., Rudd, A. G., Wang, Y., Douiri, A., Wolfe, C. D., & Bray, B. (2020). A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLOS One*, *15*(6), e0234722.

Wang, X., & Kattan, M. W. (2020). Cohort studies: Design, analysis, and reporting. *Chest*, *158*(1), S72–S78.

Wasserman, G. A., McReynolds, L. S., Lucas, C. P., Fisher, P., & Santos, L. (2002). The voice DISC-IV with incarcerated male youths: Prevalence of disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(3), 314–321.

Waters, H. R., Hyder, A. A., Rajkotia, Y., Basu, S., & Butchart, A. (2005). The costs of interpersonal violence—an international review. *Health Policy*, *73*(3), 303–315.

Waters, H. R., Hyder, A. A., Rajkotia, Y., Basu, S., Rehwinkel, J. A., & Butchart, A. (2004). *The economic dimensions of interpersonal violence.* Department of Injuries and Violence Prevention, World Health Organization. Geneva, Switzerland.

Wathen, C. N., & MacMillan, H. L. (2018). The role of integrated knowledge translation in intervention research. *Prevention Science*, *19*(3), 319–327.

Webb, P., Bain, C., & Page, A. (2017). *Essential epidemiology: An introduction for students and health professionals* (Third edition.). Cambridge University Press.

Weisburd, D., Lum, C. M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *The Annals of the American Academy of Political and Social Science*, *578*(1), 50–70.

Welsh, B. C., Peel, M. E., Farrington, D. P., Elffers, H., & Braga, A. A. (2011). Research design influence on study outcomes in crime and justice: A partial replication with public area surveillance. *Journal of Experimental Criminology*, *7*(2), 183–198.

Wessler, B. S., Nelson, J., Park, J. G., McGinnes, H., Gulati, G., Brazil, R., Van Calster, B., van Klaveren, D., Venema, E., Steyerberg, E., et al. (2021). External validations of cardiovascular clinical prediction models: A large-scale review of the literature. *Circulation: Cardiovascular Quality and Outcomes*, *14*(8), e007858.

Wexler, H. K., De Leon, G., Thomas, G., Kressel, D., & Peters, J. (1999). The Amity prison TC evaluation: Reincarceration outcomes. *Criminal Justice and Behavior*, *26*(2), 147–167.

Wexler, H. K., Melnick, G., Lowe, L., & Peters, J. (1999). Three-year reincarceration outcomes for Amity in-prison therapeutic community and aftercare in California. *The Prison Journal*, *79*(3), 321–336.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399.

Whiting, D., Gulati, G., Geddes, J. R., & Fazel, S. (2021). Association of schizophrenia spectrum disorders and violence perpetration in adults and adolescents from 15 countries: A systematic review and meta-analysis. *JAMA Psychiatry*, *79*(1), 120–132.

Whiting, D., Lichtenstein, P., & Fazel, S. (2021). Violence and mental disorders: A structured review of associations by individual diagnoses, risk factors, and risk assessment. *Lancet Psychiatry*, *8*(2), 150–161.

WHO. (2019). *Azerbaijan brings quality tuberculosis care to prisons*. Retrieved June 25, 2022, from https://www.who.int/news-room/feature-stories/detail/azerbaijan-brings-quality-tuberculosis-care-to-prisons

WHO. (2022). *Suicide rates*. Retrieved July 26, 2022, from https://www.who.int/data/gho/data/themes/mental-health/suicide-rates

Wibbelink, C. J. M., Hoeve, M., Stams, G. J. J. M., & Oort, F. J. (2017). A meta-analysis of the association between mental disorders and juvenile recidivism. *Aggression and Violent Behavior*, *33*, 78–90.

Widom, C. S., & Wilson, H. W. (2015). Intergenerational transmission of violence. In J. Lindert & I. Levav (Eds.), *Violence and mental health: Its manifold faces* (pp. 27–45). Springer.

Wildeman, C., & Wang, E. A. (2017). Mass incarceration, public health, and widening inequality in the USA. *Lancet*, *389*(10077), 1464–1474.

Willoughby, M., Young, J. T., Spittal, M. J., Borschmann, R., Janca, E., & Kinner, S. A. (2021). Violence-related deaths among people released from incarceration: Systematic review and meta-analysis of cohort studies. *EClinicalMedicine*, *41*, 101162.

Wilson, D. B., Bouffard, L. A., & MacKenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice and Behavior*, *32*(2), 172–204.

Wilson, H. W., Smith Stover, C., & Berkowitz, S. J. (2009). Research review: The relationship between childhood violence exposure and juvenile antisocial behavior: A meta-analytic review. *Journal of Child Psychology and Psychiatry*, *50*(7), 769–779.

Winetsky, D. E., Almukhamedov, O., Pulatov, D., Vezhnina, N., Dooronbekova, A., & Zhussupov, B. (2014). Prevalence, risk factors and social context of active pulmonary tuberculosis among prison inmates in Tajikistan. *PLOS One*, *9*(1), e86046.

Wishart, G. C., Azzato, E. M., Greenberg, D. C., Rashbass, J., Kearins, O., Lawrence, G., Caldas, C., & Pharoah, P. D. (2010). Predict: A new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Research*, *12*(1), 1–10.

Wolf, A., Gray, R., & Fazel, S. (2014). Violence as a public health problem: An ecological study of 169 countries. *Social Science & Medicine*, *104*, 220–227.

Wolff, N., Shi, J., Blitz, C. L., & Siegel, J. (2007). Understanding sexual victimization inside prisons: Factors that predict risk. *Criminology & Public Policy*, *6*(3), 535–564.

Wolff, N., Shi, J., & Siegel, J. A. (2009a). Patterns of victimization among male and female inmates: Evidence of an enduring legacy. *Violence and Victims*, *24*(4), 469–484.

Wolff, N., Shi, J., & Siegel, J. A. (2009b). Understanding physical victimization inside prisons: Factors that predict risk. *Justice Quarterly*, *26*(3), 445–475.

Wolff, R. F., Moons, K. G., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., Mallett, S., & PROBAST Group. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, *170*(1), 51–58.

Wood, A. M., Royston, P., & White, I. R. (2015). The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. *Biometrical Journal*, *57*(4), 614–632.

World Health Assembly, 4. (1996). *Prevention of violence: Public health priority*. Geneva, Switzerland.

World Health Organization. (2014a). *Global status report on violence prevention 2014*. Geneva, Switzerland.

World Health Organization. (2014b). *Injuries and violence: The facts 2014*. Geneva, Switzerland.

Yokotani, K., & Tamura, K. (2015). Effects of personalized feedback interventions on drug-related reoffending: A pilot study. *Prevention Science*, *16*(8), 1169–1176.

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

Yoshinaga, C., Kadomoto, I., Otani, T., Sasaki, T., & Kato, N. (2004). Prevalence of post-traumatic stress disorder in incarcerated juvenile delinquents in Japan. *Psychiatry and Clinical Neurosciences*, *58*(4), 383–388.

Young, S., Moss, D., Sedgwick, O., Fridman, M., & Hodgkins, P. (2015). A meta-analysis of the prevalence of attention deficit hyperactivity disorder in incarcerated populations. *Psychological Medicine*, *45*(2), 247–258.

Yu, R., Långström, N., Forsman, M., Sjölander, A., Fazel, S., & Molero, Y. (2022). Associations between prisons and recidivism: A nationwide longitudinal study. *PLOS One*, *17*, 1–18.

Yukhnenko, D., Sridhar, S., & Fazel, S. (2020). A systematic review of criminal recidivism rates worldwide: 3-year update. *Wellcome Open Research*, *4*(28), 28.

Yukhnenko, D., Wolf, A., Blackwood, N., & Fazel, S. (2019). Recidivism rates in individuals receiving community sentences: A systematic review. *PLOS One*, *14*(9), 1–15.

Zhong, S., Senior, M., Yu, R., Perry, A., Hawton, K., Shaw, J., & Fazel, S. (2021). Risk factors for suicide in prisons: A systematic review and meta-analysis. *Lancet Public Health*, *6*(3), e164–e174.

Zhong, S., Yu, R., Cornish, R., Wang, X., & Fazel, S. (2021). Assessment of violence risk in 440 psychiatric patients in China: Examining the feasibility and acceptability of a novel and scalable approach (FoVOx). *BMC Psychiatry*, *21*(1), 1–9.

Zhou, J., Chen, C., Wang, X., Cai, W., Zhang, S., Qiu, C., Wang, H., Luo, Y., & Fazel, S. (2012). Psychiatric disorders in adolescent boys in detention: A preliminary prevalence and case–control study in two Chinese provinces. *Journal of Forensic Psychiatry & Psychology*, *23*(5-6), 664–675.

Zlodre, J., & Fazel, S. (2012). All-cause and external mortality in released prisoners: Systematic review and meta-analysis. *American Journal of Public Health*, *102*(12), e67–e75.

Zlotnick, C., Johnson, J., & Najavits, L. M. (2009). Randomized controlled pilot study of cognitive-behavioral therapy in a sample of incarcerated women with substance use disorder and PTSD. *Behavior Therapy*, *40*(4), 325–336.