STRUCTURED PROFESSIONAL JUDGEMENT APPROACH TO RISK ASSESSMENT: GENERALISABILITY ACROSS PATIENT GROUPS FOR THE PREDICTION OF ADVERSE OUTCOMES IN SECURE MENTAL HEALTH CARE

A thesis submitted for the degree of Doctor of Philosophy (PhD) by Publication

by

Laura O'Shea

School of Social & Health Sciences,

Abertay University.

October, 2016

Declaration

Candidate's declarations:

I, Laura O'Shea, hereby certify that this thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy (PhD) by Publication, Abertay University, is wholly my own work unless otherwise referenced or acknowledged. This work has not been submitted for any other qualification at any other academic institution.

Signed [candidates signature]…………………………………………….

Date………………………………………………………………………..

Supervisor's declaration:

I, Professor Geoffrey Dickens hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy (PhD) by Publication in Abertay University and that the candidate is qualified to submit this thesis in application for that degree.

Signed [Principal Supervisors signature]…………………………………...

Date………………………………………………………………………..

Certificate of Approval

I certify that this is a true and accurate version of the thesis approved by the examiners, and that all relevant ordinance regulations have been fulfilled.

Supervisor……………………………………………………………..…….

Date………………………………………………………………………..

Acknowledgements

Dedication

To Tom,

"When the world feels full of noise my heart knows what to do, it finds that still and settled place and dances there with you" [Edward Monkton]

Thank you for never failing to make me smile.

To Beckie,

I am who I am in no small part because of you. Thank you for always believing in me and pushing me to be the best I can be.

Abstract

This thesis comprises a rigorous and coherent body of work related to the use of the HCR-20 and the START to inform risk assessment and management of secure mental health inpatients. The thesis contributes significant theoretical and applied knowledge by: 1) investigating the extent to which these tools can be generalised beyond restricted validation samples to the full range of individuals in contact with secure services, 2) determining whether they can aid assessment and management of adverse outcomes beyond aggression, and 3) offering practical, empirically-derived advice for clinicians regarding management strategies that may reduce the occurrence of adverse events. This collection of papers has used considerably novel methods, such as rocreg analysis in risk assessment of behavioural outcomes, and high quality, routinely collected data to gain a more realistic representation of what occurs in clinical practice. Further, the papers draw on larger sample sizes than have previously been reported in this area, allowing for more complex statistical analysis. This thesis has helped clarify the contexts in which these instruments perform effectively and therefore has important implications for clinical risk assessment in inpatient settings. Specifically, there is evidence that the HCR-20 and the START may aid assessment and management of aggression for the majority of groups examined, and that both tools have some efficacy for predicting self-harm among female populations. However, the HCR-20 should not be used to inform prediction and management of aggression and self-harm for individuals with developmental and organic disorders and is unlikely to be beneficial at informing risk management strategies targeting self-harm among males; the START should not be used to inform prediction and management of substance abuse, victimisation, or unauthorised leave. Finally, this thesis highlights a number of directions for future research to continue advancement in this area.

Table of Contents

List of Figures

List of Tables

Abbreviations

AUC	Area Under the [ROC] Curve
AUDIT	Alcohol Use Disorders Identification Test
DBT	Dialectical Behaviour Therapy
DOR	Diagnostic Odds Ratio
DSM-IV	Diagnostic and Statistical Manual of mental disorders 4th Revision
DUDIT	Drug Use Disorders Identification Test
FAM	Female Additional Manual
FN	False Negatives
FP	False Positives
HCR-20	Historical, Clinical, Risk-Manegement-20
ICD-10	The International Statistical Classification of Diseases and Related Health Problems 10th Revision
ID	Intellectual Disability
MOAS	Modified Overt Aggression Scale
NND	Number Needed to Detain
NPV	Negative Predictive Value
NSD	Number Safely Discharged
OAS	Overt Aggression Scale
OR	Odds Ratio
PPV	Positive Predictive Value

RAGEE	The Risk Assessment Guidelines for the Evaluation of Efficacy
RCT	Randomised Controlled Trial
ROC	Receiver Operating Characteristic
SOS	START Outcome Scale
STARD	The Standards for Reporting of Diagnostic Accuracy
START	Short-Term Assessment of Risk and Treatability
THREAT	Threats of Harm that are Real, Enactable, Acute and Targeted
TN	True Negatives
TP	True Positives

1 Introduction

1.1 Overview

The research presented in this thesis was published in peer-reviewed journals between 2013 and 2016. They are the result of a planned program of research on the use of the structured professional judgement approach to risk assessment among adults in secure mental health settings, which were conducted as two related projects: 1) research with the Historical, Clinical, Risk-Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997; Webster, Eaves, Douglas, & Wintrup, 1995), conducted in 2012-2014, and 2) research with the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009; Webster, Martin, Brink, Nicholls, & Middleton, 2004), conducted 2013-2015.

The HCR-20 is currently in its third version (Douglas, Hart, Webster, & Belfrage, 2013); however, version 3 was not available at the point of data collection and, therefore, the research presented in this thesis concerns the performance of version 2 (Webster et al., 1997). The basic domains covered by the two versions are similar but a number of items have been removed or replaced; sub-items have also been added to a number of the risk factors to both provide clarity and promote consideration of all aspects of the factor. Further, additional information has been added to indicate how risk factors might manifest (see Douglas, Hart, et al., 2014 for a comprehensive overview of changes). However, research has found that correlations between version 2 and version 3 of the HCR-20 were .91 and .98 for the total score and summary judgement respectively, suggesting individuals will score similarly on the two versions and be assigned the same level of overall risk; correlations between subscale scores ranged from .76 to .89 (Strub, Douglas, & Nicholls, 2014). Further, correlations between the two instruments and violence were equivalent at 6-8 month follow-ups (Strub et al., 2014), indicating the two versions are equally predictive of violence. Therefore, it is highly likely that the findings reported in this thesis will also apply to version 3.

This chapter describes secure mental health care internationally and in the UK, provides background information about St. Andrew's (where the data for the

primary research papers presented in this thesis was collected), demonstrates the importance of accurate and informative risk assessment, and outlines the aim of the thesis. The second chapter addresses methodological considerations for risk assessment research in secure mental health care, including: defining risk and protective factors, approaches to risk assessment, data collection, statistical methods, reporting guidelines, and ethical considerations. Chapter 3 summarises the original empirical work conducted with the HCR-20 and the START that form the basis for this thesis (full papers are presented as appendices), the rationale behind these papers, and the contributions the candidate made to these. The final chapter summarises the novel contributions that the thesis has made to the literature, the implications these have for the use of the HCR-20 and the START in secure mental health settings, and presents future directions for research. This thesis aims to demonstrate that the candidate has made a substantial contribution to the empirical work outlined above, that such work has contributed significant new knowledge and has implications for applied risk assessment and management in secure mental health populations, and that the work presented has had a substantial impact, as demonstrated by the number of citations and presentations at national and international conferences (see 5.1).

1.2 Secure mental health care

The large numbers of mentally disordered offenders detained in prisons worldwide highlights the need for secure mental health care (Sugarman & Dickens, 2015). A recent survey of World Health Organisation member states, covering 98% of the world's population, found that 80% of countries have dedicated secure mental health hospitals (Morris, Lora, McBain, & Saxena, 2012). The purpose of these hospitals is to provide treatment for adults with mental disorder that are at significant risk to themselves or others (Centre for Mental Health, 2011). There is limited information available about the configuration of secure mental health services from a global perspective (for an overview, see Sugarman & Dickens, 2015). However, there is wide variation in the number of beds available, with Europe having 35 beds per 100,000 of the population and all other regions having less than 10 beds per 100,000 of the population (World Health Organisation, 2014). This may reflect the view held by

a number of European countries, including the United Kingdom, that it may be more suitable to provide mental health treatment for offenders outside of prisons. In contrast, in the USA the odds of an individual with serious mental illness being in prison compared with hospital is 3.2 (Torrey, Kennard, Eslinger, Lamb, & Pavle, 2010); this ranges from less than 2.0 in Connecticut, Maine, Massachusetts, Minnesota, Mississippi, Nebraska, New Jersey, New York, North Dakota, Rhode Island, and Vermont to greater than 7.0 in Arizona, Nevada, and Texas. In recent years, there has been a reduction worldwide in the number of beds available with a 28% reduction in figures between the 2011 and 2014 Mental Health Atlases (World Health Organisation, 2011, 2014); the greatest reduction was for upper-middle income countries (44%) with low income countries showing a 16% reduction. These results may in part reflect deinstitutionalisation: the ongoing shift away from long-term hospitalisation in favour of increased mental health services in the community (Csipke, Flach, McCrone, Rose, & Tilley, 2014).

Secure services typically provide care across varying security levels to reflect the risks and needs of the broad range of patients who present to them (Fisher, Geller, & Pandiani, 2009). In England and Wales, all high secure services are provided by the NHS and are aimed at individuals who are detained under the Mental Health Act (1983, amended 2007) and pose a "grave and immediate danger to the public" (ibid. p.12). Medium secure services are designed for individuals who are similarly detained under the Mental Health Act and pose a serious risk to the public; approximately 65% of beds are provided by the NHS, the remainder are provided by the independent sector (i.e., for profit and not for profit organisations that work in partnership with the NHS and are bound by the NHS constitution; Department of Health, 2015). Low secure services, also provided by both the NHS and independent sector, are less well defined but provide beds for individuals detained under the Mental Health Act who "pose significant danger to themselves or others" (ibid. p.13). Forensic services are contained within the wider secure context and are concerned with individuals who have been admitted directly to hospital, or transferred from prison to secure services, following a criminal conviction in order to have access to treatment and care for a mental disorder (Centre for Mental Health, 2011).

1.2.1 St. Andrew's

All of the original empirical data for the papers presented in this thesis was collected at St. Andrew's (http://www.stah.org/about-st-andrews), a large charitable provider of inpatient mental health care with over 900 beds located at four sites in England. Accommodation is gender specific and provided in medium secure, low secure, locked, open rehabilitation wards, and psychiatric intensive care. Currently, there are six secure care pathways: men's mental health, women's mental health, adolescent, learning disability, autistic spectrum disorder, and neuropsychiatry (acquired brain injury, Huntington's disease, and dementia). The population comprises those detained under a forensic commitment (i.e., the individual was deemed sufficiently unwell to require hospitalisation at the time of sentencing or was transferred from prison to hospital for treatment), those detained under a civil commitment (i.e., detained for their own or other's safety in the absence of a conviction), and those admitted voluntarily.

Examination of the number of beds available in each service (St Andrew's Healthcare, 2016) indicates some similarities and differences compared with the wider secure mental health population in the UK. In 2014, 66% of individuals detained under the Mental Health Act in England were male (Health & Social Care Information Centre, 2014); 58% of beds at St. Andrew's are dedicated to males, but an additional 9% of available beds are on mixed wards (St Andrew's Healthcare, 2016). The proportion of beds available specifically for individuals with personality disorder at St. Andrew's (13%; St Andrew's Healthcare, 2016) was also similar to the proportion of individuals with personality disorder that have been discharged from medium secure units (14%; Coid, Hickey, Kahtan, Zhang, & Yang, 2007). However, Coid et al. (2007) reported that only 5% of individuals discharged from medium secure units had an organic disorder, whereas 20% of the beds at St. Andrew's are for neuropsychiatric disorders (St Andrew's Healthcare, 2016). As the aforementioned study is concerned with discharged individuals, rather than those currently in care, it is possible that observed differences are in part due to lower rates of discharge among patients with organic disorders; however, discharged individuals accounted for a large proportion (83%) of the total sample. Finally, there is a lack of available

information regarding the proportion of individuals with neurodevelopmental disorders in secure mental health care; however, such beds account for 24% of those available at St. Andrew's (St Andrew's Healthcare, 2016) which is likely to be much higher than the rate found in secure services more widely due to the highly specialist nature of these services. Therefore, the unique nature of this population, and the fact that the HCR-20 and the START are now mandated in many secure services including St. Andrew's, highlights the need to investigate the performance of these tools among the full range of individuals coming into contact with secure services.

1.3 The importance of accurate risk assessments

Risk assessment is a vital task for mental health professionals and influences a variety of decisions regarding sentencing, admission to and discharge from secure mental health hospitals, required security level, and, more recently, treatment and management plans (Dvoskin & Heilbrun, 2001; Flynn, O'Neill, McInerney, & Kennedy, 2011; Skeem et al., 2005). Errors in risk assessment are costly; when a patient is incorrectly deemed at risk (i.e., false positive error), consequences for that individual may include unnecessary restriction of civil liberties (de Ruiter & Nicholls, 2011; Szmukler & Rose, 2013) such as involuntary commitment and forced use of medication, damage to the therapeutic relationship (Granello, 2010; Skeem et al., 2005), and increased stigma (Large, Ryan, Singh, Paton, & Nielssen, 2011). False positive errors also lead to mismanagement of limited resources (de Ruiter & Nicholls, 2011) and is financially costly; the average cost of a bed in a medium secure mental health hospital has been estimated at over £200,000 per year (Freestone et al., 2012). In contrast, false negative errors, when a high risk patient is deemed low risk, may result in the occurrence of potentially preventable adverse events (Large et al., 2011) such as harm to the patient themselves, family members, or members of the general public (Skeem et al., 2005).

There are high rates of challenging behaviour in secure mental health services with approximately 80% of nurses experiencing violence during their career (Bowers et al., 2011). More specifically, Nijman, Bowers, Oud and Jansen (2005) reported that 89% of psychiatric nurses witnessed verbal aggression,

84% self-harm, 76% mild physical violence, 68% a suicide attempt, 28% a completed suicide, and 16% severe physical violence over a one-year period. Experiencing such events is associated with days absent from work for illness (James, Stewart, & Bowers, 2012; Lanctôt & Guay, 2014; Nijman et al., 2005), damage to the therapeutic relationship (Lanctôt & Guay, 2014), mental health problems such as PTSD and anxiety (Lanctôt & Guay, 2014), and therapeutic nihilism (James et al., 2012). Unauthorised leave is another behaviour that causes concern among mental health professionals; further, its occurrence is associated with an increase in the rates of other challenging behaviours such as self-harm, suicidal behaviours, self-neglect, and aggression towards others (Bowers, Jarrett, & Clark, 1998; Winship, 2011). More specifically, numerous sources have reported that approximately one quarter of inpatient suicides occur when the individual is absent without staff agreement (Bowers et al., 1998; National Confidential Enquiry into Suicide and Homicide by People with Mental Illness, 2014). Therefore, determining the efficacy of standardised risk assessment tools that can assist mental health professionals in managing such behaviours is crucial to maintaining patient and staff safety and promoting a therapeutic environment.

1.3.1 Consideration of sample characteristics

One area that has drawn a reasonable amount of attention, at least theoretically if not in empirical investigations, is gender differences in risk. Proponents of the "gendered perspective" (Nicholls, Ogloff, & Douglas, 2004) argue that crime and violence perpetrated by women is closely linked with their unique experiences as women. For example, higher rates of childhood abuse in women compared with men may lead to behaviours such as running away from home (Chesney-Lind, 1989). Once on the streets women may turn to crime in order to survive (McCormack, Janus, & Burgess, 1986). Further, the higher rates of common mental disorders among women compared with men (Mental Health Foundation, 2015) may increase the likelihood of them engaging in adverse outcomes.

The argument that our demographic characteristics shape the type of risks we encounter, and how we respond to such risks, should apply equally to other risk

outcomes and demographic characteristics. For example, it has been suggested that social inequality and racial discrimination experienced by ethnic minorities leads to social isolation, restricted socioeconomic mobility, structural barriers, and stress, that may increase risk of developing mental health problems and engaging in violent or criminal behaviour (Sampson & Wilson, 2005; Williams & Williams-Morris, 2000). Further, threat/control-override symptoms, defined as delusions that others are out to harm the individual, that the individual is dominated by external forces, or experiencing thought insertion (Nolan et al., 2005), are linked to elevated rates of violence among patients with schizophrenia (Nederlof, Muris, & Hovens, 2011); although this has not been consistently found (Bo, Abu-Akel, Kongerslev, Haahr, & Simonsen, 2011).

Differences in risk factors for suicide have been found across a number of demographic factors. For example, having a young child appears to a protective factor against suicide for women only, whereas unemployment, retirement, and sickness-related absence were risk factors for suicide only among men (Qin, Mortensen, Agerbo, Westergard-Nielsen, & Eriksson, 2000). Similarly, Kung, Liu and Juon (1998) found that having a higher degree of education, living alone, previous use of mental health services, heavy alcohol consumption, and having a "blue-collar" occupation (i.e., jobs involving manual labour) were all associated with increased risk of suicide among Caucasians, whereas only previous use of mental health services was associated with increased risk in African-Americans. Further, being well educated, having a higher IQ, having high personal expectations, fear of further mental deterioration, and loss of faith in, or high dependence on, treatment appear to be specific risk factors for suicidal behaviour in schizophrenia (Caldwell & Gottesman, 1990; Hawton, Sutton, Haw, Sinclair, & Deeks, 2005; Siris, 2001). However, impulsivity and childhood abuse appear to be more important factors for suicide risk in borderline personality disorder (Black, Blum, Pfohl, & Hale, 2004; Soloff, Lynch, & Kelly, 2002). Demographic characteristics may also interact with one another; for example, depressive symptoms are only associated with suicide in older men, whilst they are a risk factor for women across all age groups (Kung, Pearson, & Liu, 2003).

Older age is a known risk factor for self-neglect (Burnett et al., 2007); however, it is unclear whether this is due to age-specific risk factors or just a greater prevalence of risk factors for self-neglect among older adults. For example, cognitive and physical impairments and living alone increase the likelihood of self-neglect (Abrams, Lachs, McAvay, Keohane, & Bruce, 2002; Iris, Ridings, & Conrad, 2010) and are more common in older adults (Office for National Statistics, 2011; Poythress, Burnett, Naik, Pickens, & Dyer, 2006). However, cognitive impairment is a serious obstacle to adequate self-care regardless of age (Choi, Kim, & Asseff, 2009). Research into self-neglect has been dominated by elder self-neglect, due to the much greater incidence of such behaviour among this population; research into self-neglect in younger adults has been limited to behaviours that occur in the context of self-harm, eating disorders, and substance abuse (Iris et al., 2010).

1.4 Aim of the thesis

Section 1.3 outlined the severe consequences of inaccurate risk assessments and possible differences in risk factors across clinically and demographically defined groups were presented in section 1.3.1. In light of this, the performance of risk assessment tools based on such factors should be examined separately for different groups in contact with mental health and criminal justice populations; this is necessary in order to determine how accurately risk assessment tools perform in the full range of groups to which they are applied. Perhaps surprisingly, relatively little research has addressed this and studies regarding the predictive efficacy of risk assessment tools have been conducted in primarily male samples. For example, a recent, large scale meta-analysis of nine risk assessment tools (N=25,980) found that 91% of samples included comprised over 50% male participants (Singh, Grann, & Fazel, 2011). The majority of studies included in this analysis did not report diagnostic or ethnic characteristics of samples; however, the higher the mean age of the sample, and the higher the proportion of Caucasian individuals, the higher the diagnostic odds ratio (DORs), suggesting superior performance. Further, there is some evidence that risk assessment tools perform more accurately in samples with a higher proportion of women (Singh et al., 2011; Yang, Wong, & Coid, 2010).

The very limited evidence available suggests that there may be some differences in the performance of risk assessment schemes across groups. Therefore, the overall aim of this thesis was to investigate the ability of the HCR-20 and the START to inform risk prediction and planning for a wide range of adverse outcomes among diverse groups of secure mental health inpatients. The purpose of this is to 1) inform risk assessment practice in secure care and minimise the use of risk assessment tools, on which restrictive decisions are often based, among patient groups for which they have not demonstrated predictive efficacy, and 2) identify those items with the greatest predictive efficacy for specific patient groups to highlight factors that will likely reduce risk if successfully targeted by interventions. Figure 1 provides an overview of the thesis and demonstrates how the presented papers link together and collectively inform risk assessment practice. More specifically, it indicates the two key research questions the thesis aimed to address, how the thesis evolved, and how new studies were devised in response to the emerging evidence.

Background: risk assessment is a crucial task for mental health professionals in secure psychiatric settings. However, risk assessment tools have not been validated in the full range of patient groups that are managed in such services. It is important to determine how effectively such tools perform, and therefore the appropriateness of their use, in the full range of patient groups to which they are applied. Errors in risk assessment can have severe consequences. False positives may lead to unnecessary restrictions and mismanagement of limited resources; false negatives may result in the occurrence of preventable negative events, causing harm to the patient themselves, staff members, other patients, or members of the public.

The HCR-20 is widely used and extensive research has demonstrated its predictive efficacy in primarily male samples. The START has received attention due to its unique features but little empirical research of these has been conducted.

RQ1: How does the predictive efficacy of the HCR-20 vary based on demographic or clinical characteristics and what are the implications of this for treatment planning?

Overall aim: to investigate the ability of the HCR-20 and the START to inform risk prediction and treatment planning for a wide range of adverse outcomes among diverse groups of secure psychiatric inpatients?

RQ2: Does the START have predictive efficacy for its intended outcomes, other than aggression and self-harm and how does the predictive efficacy vary based on demographic and clinical characteristics; what are the implications of this for treatment planning?

Rationale: Previous reviews of the HCR-20 have not examined moderators of predictive efficacy. Important to determine if it performs effectively in full range of patients.

Rationale: No previous systematic review. Important to determine if it is an improvement over more established tools, such as the HCR-20

Paper 1: Moderators of the predictive efficacy of the Historical, Clinical and Risk Management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis

Rationale: Some evidence that effect sizes differ based on gender, diagnosis, and ethnicity breakdown of studies included in paper 1. Should directly examine the impact of these on predictive efficacy

Rationale: Paper 2 revealed no evidence the START predicts outcomes beyond aggression and self-harm but based on small number of studies

Paper 2: START: Systematic review and meta-analysis

Paper 3: Differential predictive validity of the HCR-20 for inpatient aggression

Paper 4: Predictive validity of the HCR-20 for inpatient aggression: the effect of intellectual disability on accuracy

Paper 6: Predictive validity of the START for unauthorised leave and substance abuse in a secure mental health setting: A pseudo-prospective cohort study

Paper 5: Insufficiently powered to examine unauthorised leave and substance abuse due to low base rates

Paper 5: The predictive validity of the START for multiple adverse outcomes in a secure psychiatric inpatient setting

Rationale: No previous research on gender differences in the performance of the START.. However, paper 1 and paper 3 found evidence that the HCR-20 has superior efficacy in women; therefore, the START may be similarly affected.

Paper 7: Predictive Validity of the START for Aggression and Self-Harm in a Secure Mental Health Service: Gender Differences

Paper 8: Predictive validity of the HCR-20 for inpatient self-harm

Rationale: Papers 2, 5 and 6 revealed limited evidence that the START predicts outcomes beyond aggression and self-harm. Only START risk estimates predict self-harm Therefore, pragmatic to assess if the HCR-20 can predict self-harm as the tool is more established.

Contribution of presented HCR-20 research to evidence based practice

The HCR-20 can inform risk prediction and treatment planning for aggressive behaviours in: 1) men and women, with superior efficacy among women, 2) those with and without intellectual disability, 3) those with schizophrenia and personality disorder, 4) Caucasians and non-Caucasians, and 5) older and younger individuals. It may be useful at informing risk of verbal aggression, but not physical aggression in those with developmental and organic disorders; therefore, clinicians may want to consider additional factors to inform risk of physical aggression in these populations.

The HCR-20 can inform risk prediction and treatment planning for self-harm in: 1) women, and 2) those with schizophrenia and personality disorder. It does not inform risk of self-harm in: 1) men, and 2) those with developmental and organic disorders.

Clinicians should consider using interventions that target impulsivity and non-compliance with remediation attempts for individual at risk of aggression and self-harm. Additionally, targeting active symptoms of major mental illness, exposure to destabilisers, stress and personal support may be beneficial for reducing aggression; for self-harm targeting negative attitudes and plans lacking feasibility are also likely to be fruitful.

Contribution of presented START research to evidence based practice

The Strength scale can inform risk prediction and treatment planning of aggression in women; the Vulnerability scale can inform risk prediction and treatment planning of aggression in men, but only verbal aggression in women. Clinicians should consider targeting impulsivity, current mental state, adherence and external triggers to reduce aggressive incidents.

The Vulnerability scale can inform risk prediction and treatment planning of self-harm in women, but not in men; therefore, clinicians should consider additional factors when judging risk of self-harm among males. Targeting adherence, coping skills, and impulse control may help reduce incidents of self-harm.

The Strength scale can inform risk prediction of self-neglect.

The START is unlikely to be useful at informing risk prediction and treatment planning for substance abuse, victimisation, or unauthorised leave; however, clinicians should consider using interventions that target substance abuse directly, and self-care, to reduce substance abuse.
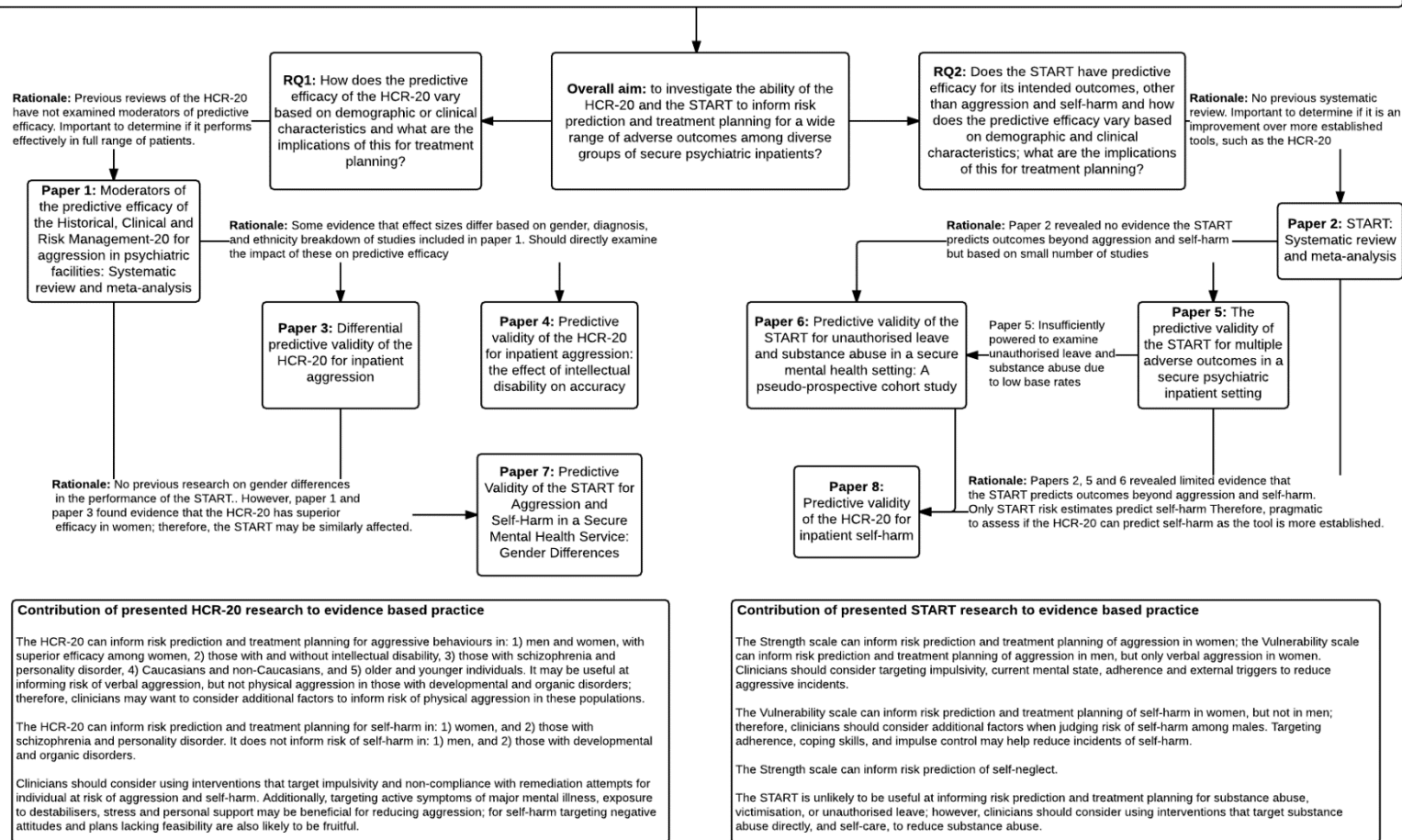
Figure 1: Overview of the relationship between papers contained in this thesis.

2   Methodology for risk assessment research in secure mental health settings

2.1   Defining risk and protective factors

The meaning of risk is widely understood; risk factors are factors that are associated with an increase in the likelihood of a negative outcome occurring (Jessor, Van Den Bos, Vanderryn, Costa, & Turbin, 1995). In the secure mental health setting, we are most commonly concerned with risk factors for violence; for example, it is widely agreed that psychopathy, previous violence, and major mental disorder are associated with an increase in risk for violence to some extent (e.g., Hart, 1998a; Monahan, 1992). In contrast, protective factors are those that are associated with a decrease in the likelihood of a negative outcome occurring (Jessor et al., 1995). This may occur by altering exposure to the risk, reducing the likelihood of negative chains of events occurring as a result of encountering a risk, promoting self-esteem, and predisposing the individual towards positive opportunities (Rutter, 1993).

There is some debate about how protective factors correspond to risk factors. Protective factors have been variously defined as the absence of a risk factor (e.g., no previous violence), a factor that lies at the opposite end of a continuum to a risk factor (e.g., history of prosocial behaviour as opposed to history of violence measured on a single continuum), or conceptually distinct, with no corresponding risk factor (e.g., history of pro-social behaviour irrespective of history of violence) (O'Shea & Dickens, 2014). The latter perspective allows an individual to possess risk and protective factors simultaneously in a given domain (Braithwaite, Charette, Crocker, & Reyes, 2010; Webster et al., 2009).

Kraemer et al. (1997) outlined a typology for risk and protective factors to increase consistency across risk-related research; for ease they refer only to risk factors but note that terminologies apply equally well to protective factors. They assert that the label we give to a factor should reflect the state of current scientific knowledge regarding that factor; a factor's label may therefore change as new information is discovered. They define a risk factor as "a measurable characterization of each subject in a specified population that precedes the

outcome of interest and can be used to divide the population into two groups [the high-risk and the low-risk groups that comprise the total population]" (ibid. p. 338). They stress the importance of presenting a measure of potency when claiming a characteristic to be a risk factor, defined as the maximum discrepancy between low and high risk achievable by that factor. Further, time is an important consideration when defining risk factors. A factor that is associated with the outcome without established precedence is a correlate of that outcome; a factor can only be labelled as a risk factor if it is shown to precede the outcome. Risk factors can then be subdivided into fixed/static markers that cannot change and variable/dynamic risk factors that can change. Variable risk factors are further divided into causal risk factors that can be manipulated and such manipulation changes the risk of the outcome occurring, and variable markers that cannot be manipulated, or manipulation produces no change in risk (Kraemer et al., 1997). The term multivariate risk factor is used to refer to situations where multiple factors are used to identify high and low risk subjects (Kraemer et al., 1997), as is theoretically the case for many risk assessment schemes. Whether such schemes actually meet the criteria for a multivariate causal risk factor has important implications for their ability to inform risk management strategies and track treatment progress.

2.2 Defining and measuring adverse outcomes

In order to accurately assess and communicate risk it is essential to clearly define the potentially adverse outcomes of interest and measure them in a reliable and valid manner (Kraemer et al., 1997); inadequate or inconsistent definitions of outcomes will likely impede research (Harris, Oakley, & Picchioni, 2013). In secure mental health settings, risk of aggression or violence is most frequently considered; however, there are a number of other adverse outcomes (e.g., self-harm, victimisation, and unauthorised leave) that may act as a barrier to successful treatment and recovery (Webster et al., 2009), which should therefore be considered (see Table 1 for measures used in this thesis).

2.2.1 Aggression and violence

There are a number of different available definitions of violence and aggression and the two terms are often used interchangeably despite having slightly

different meanings, with violence being a more serious form of aggression (Harris et al., 2013). Aggression may be defined as "behavior directed toward another individual that is carried out with the proximate (immediate) intent to cause harm. In addition, the perpetrator must believe that the behavior will harm the target, and that the target is motivated to avoid the behavior" (Anderson & Bushman, 2002; p. 28). Violence is aggression that is intended to cause severe harm; therefore, all violence can be considered as aggression but not all aggression would be considered violent (Anderson & Bushman, 2002). For example, the World Health Organisation define violence 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, Mercy, Dahlberg, & Zwi, 2002; p. 1084).

Violence and aggression can be seen as having four central components that should be considered when measuring their occurrence: planning, intent, nature, and outcome/consequences (Harris et al., 2013). The definitions presented above revolve around intent; however, few attempts have been made to develop standardised outcome assessment tools that incorporate intent. One exception to this is the Quantification of Violence Scale (Tyrer et al., 2007), which aims to asses planning, intent and consequences. However, perceived intent was found to have the lowest inter-rater reliability, reflecting difficulties in accurately establishing this component. The Attempted and Actual Assault Scale (Attacks; Bowers, Nijman, & Palmstierna, 2007) measures commitment to do harm, which may be seen as similar to intent. However, other outcome measures are primarily concerned with the nature of the act or the consequences to the victim (i.e., injury).

Among the most widely used scales are the MacArthur Community Violence Instrument (Steadman et al., 1998) and the Modified Overt Aggression Scale (MOAS; Kay, Wolkenfeld, & Murrill, 1988). The MacArthur Community Violence Instrument assesses the nature of aggression by examining the occurrence and frequency of the following categories of aggression, typically using self-report and collateral information: throwing an object; pushing, grabbing or shoving; slapping; kicking; biting; choking; hitting or beating up; forcing sex; threats with

a weapon; and use of a weapon (Steadman et al., 1998). The original study retrospectively assessed behaviours occurring in the last 10 weeks but the tool has since been used by a number of studies to assess history of aggression and violence over longer time periods (Harris et al., 2013). For example, the MacArthur Community Violence Instrument has been used to assess occurrence of such behaviours over six months (e.g., Johnson, Desmarais, Van Dorn, & Grimm, 2015; Swanson, Van Dorn, Monahan, & Swartz, 2006) and across an individual's lifetime (e.g., McGregor, Castle, & Dolan, 2012; Tsigebrhan, Shibre, Medhin, Fekadu, & Hanlon, 2014); however, the reliability of this instrument over such time periods is scarcely reported. The MOAS, and the Overt Aggression Scale (Yudofsky, Silver, Jackson, Endicott, & Williams, 1986) from which it was modified, assess four categories of aggressive behaviour: verbal aggression, physical aggression against objects, physical aggression against self, and physical aggression against others. Each category is rated on a five-point scale, from 0 (no incident of that nature) to 4 (most severe incident); for example, the categories of verbal aggression are: 0=no verbal aggression; 1=makes loud noises, shouts angrily; 2=yells mild personal insults, (e.g., "You're stupid!"); 3=curses viciously, uses foul language in anger, makes moderate threats to others or self; and 4=makes clear threats of violence toward others or self (e.g., "I'm going to kill you") or requests help to control self (Kay et al., 1988; Yudofsky et al., 1986). To score above a two on the physical aggression against self and others categories, injury needs to be apparent; therefore, this scale can be seen as assessing the nature and outcome of aggression. The MOAS provides the addition of a weighting system, such that severity scores are multiplied based on the category of aggression they belong to (Verbal aggression=1, physical aggression against objects=2, physical aggression against self=3, and physical aggression against others=4) (Kay et al., 1988); the higher the resulting score, the more serious the incident.

For papers 3, 4, and 8 of this thesis, members of the research team coded aggressive incidents according to the OAS/MOAS criteria (see 5.2 for coding sheet). The weighting from the MOAS was used to determine the most serious outcome when electronic progress notes contained more than one adverse incident but weighted scores were not reported in the papers. This decision was

made for a number of reasons: 1) as the OAS/MOAS is one of the most commonly used tools and has a strong evidence base (Harris et al., 2013), 2) the MacArthur Community Violence Instrument is intended for use in the community and contains categories of behaviour that may be extremely rare in a secure mental health setting (Steadman et al., 1998), and 3) the presented papers used routinely collected data from which it would be difficult to establish intent and planning. The START Outcome Scale (SOS; Nicholls et al., 2007; see 5.3 for coding sheet) was used to code aggressive incidents for papers 5 and 8. The SOS is modified from the MOAS and measures the occurrence of 12 categories of adverse behaviour, including verbal aggression, aggression against property, and physical aggression against others as defined in the OAS/MOAS. The SOS was used instead of the OAS/MOAS as these papers presented data regarding multiple adverse outcomes and the use of a single measure to capture these is more pragmatic; however, the tools are essentially equivalent for incidents of aggression and self-harm.

2.2.2 Self-harm and suicidal behaviour

Suicide can be clearly defined as "the act of deliberately killing oneself" (World Health Organization, 2014; p. 12) and the term suicidal behaviour can be used to encompasses ideation, planning suicide, attempting suicide, and suicide itself (ibid). However, terminology and definitions for other auto-destructive acts that do not result in death is inconsistent, although self-harm is generally preferred (Fliege, Lee, Grimm, & Klapp, 2009) and will be used throughout this thesis. Self-harm can be defined as "the deliberate, direct destruction or alteration of body tissue without conscious suicidal intent, but resulting in injury severe enough for tissue damage to occur" (Gratz, 2001; p. 253); however, the term is often used more broadly to encompass other self-harming behaviours without visible injury, such as pulling hair, self-hitting, exercising to hurt oneself, and stopping medication or starving with intent to cause harm (Skegg, 2005). Fliege et al. (2009) reports that the following are generally not categorised as self-harm: behaviours that are symptoms of other disorders (e.g., eating disorders), everyday unhealthy behaviours such as lack of exercise and overeating, and psychological self-harm, such as deliberately engaging in an abusive relationship.

There have been a number of attempts to developed standardised tools to assess the occurrence of self-harm. The Deliberate Self-Harm Inventory (Gratz, 2001) is a 17-item self-report measure that asks about the occurrence, frequency, duration, and severity of 16 self-harming behaviours, with an additional question regarding behaviours not covered by the questionnaire. The OAS/MOAS (Kay et al., 1988; Yudofsky et al., 1986) rates self-harm on 5 severity levels: 0=no physical aggression against self, 1=picks or scratches skin, hits self, pulls hair (with no or minor injury only), 2=bangs head, hits fist into objects, throws self onto floor or into objects (hurts self without serious injury), 3=small cuts or bruises, minor burns, and 4=Mutilates self, causes deep cuts, bites that bleed, internal injury, fracture, loss of consciousness, loss of teeth. The SOS (Nicholls et al., 2007) contains self-harm as defined in the OAS/MOAS, with the additional categories of suicide ideation and planning and suicide behaviours. Suicide ideation and planning ranges from 1='occasionally feels life isn't worth living. Abstract and rare thoughts of suicide without intent or plans and no associated distress', to 4='specific detailed plan and intent. Searches for appropriate means and time; plan reflects low likelihood of resuscitation or discovery, readily available means'. Suicide behaviours ranges from 1='makes impulsive gesture or attempts with low risk of lethality', to 4='serious attempt with low likelihood of resuscitation or discovery, or completed suicide'.

For paper 8, the OAS/MOAS criteria was used to code incidents of self-harm in the same way as aggression and violence were coded; however, the SOS was used for papers 5 and 7 as this includes suicidal behaviours in addition to self-harm and defines these outcomes as intended by the START's authors (Webster et al., 2009). The decision was made to combine self-harm, suicide ideation and planning, and suicide behaviours into a composite self-harm/suicide category for both these papers due to: 1) the low rate of completed suicides in mental health inpatients (0.06-5.66 per 1000 admissions; Bowers, Banda, & Nijman, 2010), and 2) the difficulty in disentangling suicide attempts from self-harm without suicidal intent (Gray et al., 2011), especially when dealing with retrospectively collected data. Further, separating self-harm and suicide (e.g., through the use of terms such as non-suicidal self-injury) may

be problematic when individuals who self-harm are at increased risk of future suicide and would be considered a lower priority for treatment (Kapur, Cooper, O'Connor, & Hawton, 2013). In fact, the definition of suicide in the START manual (Webster et al., 2009) includes self-injurious behaviour defined as "all behaviours that involve deliberate infliction of direct physical harm to one's body with zero intent to die as a consequence of this behaviour" (ibid, p. 13); therefore, there is an overlap between self-harm and suicide as defined by the START authors. If clinicians are following available guidance and forming specific risk estimates for suicide using the definition provided, then the composite outcome used is consistent with this definition.

2.2.3 Self-neglect

Self-neglect can be considered as the inability to provide the self with the things required to meet basic needs (Dyer, Goodwin, Pickens-Pace, Burnett, & Kelly, 2007). However, major challenges exist in consistently defining self-neglect, in part because social norms, cultural norms and changes in context (e.g., presence of caregivers) can affect what behaviours may be classified as self-neglect (O'Brien, Thibault, Turner, & Laird-Fick, 2000). NANDA international suggest two subcategories of self-neglect: primary or intentional self-neglect that can be seen as a lifestyle choice that is not indicative of an underlying disorder, and secondary or non-intentional self-neglect which is a manifestation of a psychiatric condition that impairs decision making capacity, such as dementia (Gibbons, Lauder, & Ludwick, 2006). However, most individuals could be considered to engage in some behaviour that could be seen as intentional self-neglect, such as being over- or under-weight, ingesting harmful substances such as alcohol and nicotine, and not complying with general medical recommendations (O'Brien et al., 2000). The difficulty is determining the threshold at which such behaviour constitutes self-neglect that is qualitatively or quantitatively different from that found in the general population and therefore requires intervention.

Considerably more attention has been paid to the development of validated assessment tools for elder abuse than for self-neglect specifically (for review, see Fulmer, Guadagno, Dyer, & Connolly, 2004). One tool that may be

considered useful for evaluating self-neglect is the Kohlman Evaluation of Living Skills (KELS; Kohlman-Thompson, 1992), which aims to identify those that have insufficient skills to live independently. However, whilst this measure was developed for use with mental health inpatients (Zimnavoda, Weinblatt, & Katz, 2002), it relies on self-report, observation, and functional assessments and hence cannot be completed without dedicated time and resources. It may not, therefore, be suitable in situations where individuals require assessments for risk of engaging in multiple adverse behaviours. The SOS contains objective ratings for self-neglect rated as 0='none', 1='mild problems in one or two domains; hygiene, sleep, diet or personal living space are somewhat below social standards; no serious implications, responds to direction', 2='many or moderate problems resulting in some negative outcomes (e.g., results in social stigma), does not respond to direction. Requires persistent prompting', 3='unsafe behaviour likely to result in serious implications (e.g., not following up on medical testing, not filling prescriptions)', and 4='demonstrates potentially life threatening behaviour (e.g., hunger strikes, not seeking emergency medical treatment)'. In paper 5 self-neglect was assessed using the SOS as: 1) the paper was intending to assess discrete incidents of self-neglect as opposed to global impairment that is captured by tools such as the KELS, and 2) the paper was assessing the occurrence of multiple adverse outcomes that are within the scope of the SOS.

2.2.4 Victimisation

Victimisation may be problematic to define; some suggest it is merely being the victim of aggressive behaviour (Hodgins, Alderton, Cree, Aboud, & Mak, 2007). Some liken it more to bullying, which can be seen as having the following components: repetitive, unprovoked, intended to cause harm, imbalance of power, and contains physical, psychological or verbal abuse (Ireland, 2005). However, these criteria, which were developed with the school environment in mind, have not been universally accepted with only repetition and imbalance of power being consistently applied (Ireland, 2005). Further, an investigation into the perception of bullying by offenders revealed that the majority of offenders thought that a single act could constitute bullying, that there is only sometimes

an imbalance of power, and that victims may have sometimes provoked their aggressor (Ireland & Ireland, 2003).

Victimisation is classified by the SOS as 0='none', 1='bullied or intimidated by others on occasion; mild emotional, psychological, financial injury, property damage, fear or intimidation', 2='abuse or verbal threats results in moderate/severe emotional injury, fear/intimidation, financial harm, but without physical injury', 3='physical assault results in mild-moderate physical injury (e.g., bruises, sprains, or welts) or non-consensual sexual touching or fondling', and 4='physical assault results in severe physical injury (e.g., broken bones, deep lacerations, internal injuries); or violent or coercive sexual assaults' (Nicholls et al., 2007). Therefore, at the more severe levels, there is a high degree of overlap with behaviours that may be classified as victimisation, aggression and violence, or inappropriate sexual behaviour. The SOS was used to measure incidents of victimisation in study 5 as it is consistent with the definition of victimisation used by the START authors, it was assessed as part of a paper assessing other outcomes covered by the SOS, and there is a lack of alternative validated measures available.

2.2.5 Unauthorised leave

Unauthorised leave, or absconding, is best defined as when an individual leaves a secure mental health setting without staff authorisation or having been discharged, or when patients on an authorised leave of absence fail to return within the agreed time (Arbee, 2014). However, there has been some variation in the literature on the length of absence before which an individual is classified as an absconder, with some suggesting applying a cut-off of 24 hours (Bowers et al., 1998; Winship, 2011). Further, in some studies action is only taken, or an individual is only classified as an absconder, if the individual was formally detained (Arbee, 2014), or if staff were concerned about the safety of the individual or others as a result of the absconsion (Bowers et al., 1998; Winship, 2011). Perhaps even more difficult to classify are those individuals who attempt but fail to abscond from secure services. The SOS (Nicholls et al., 2007) classifies actual and attempted unauthorised leave on a scale of 0-5; 0='none', 1='returns late from unescorted leave without prior notification or adequate

explanation', 2='returns from unescorted leave 24 hours or more late', 3='absconds from escorted leave or is returned by police from unescorted leave or does not return, Attempted escape from secure setting', and 4='escapes from secure setting'. Whilst most of these can be thought of as objective, there may be some difficulty in determining what constitutes attempted escape. The SOS was used to code incidents of unauthorised leave for paper 6 as this outcome was assessed as part of a wider study on the START that included other adverse outcomes covered by the SOS and there is a lack of alternative standardised measures.

2.2.6 Substance abuse

The International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10; World Health Organisation, 1992) classifies harmful substance use as pattern of use that is damaging to physical (e.g., hepatitis from injected substances) or mental (e.g., depression secondary to heavy alcohol consumption) health. Substance dependence is defined as behavioural, cognitive, and physiological phenomena developing from repeated substance use, typically involving factors such as a strong desire to take the drug, increased tolerance, and sometimes withdrawal (World Health Organisation, 1992).

Numerous tools are available to assess behaviours related to substance abuse, mainly via self-report, such as the 10 item Alcohol Use Disorders Identification Test (AUDIT; Saunders, Aasland, Babor, de la Fuente, & Grant, 1993) and the 11 item Drug Use Disorders Identification Test (DUDIT; Berman, Bergman, Palmstierna, & Schlyter, 2005). However, these tools are used as screening instruments to identify those with likely substance abuse disorders, demonstrating acceptable sensitivity and specificity for identifying substance abuse against ICD-10 (World Health Organisation, 1992) and the Diagnostic and Statistical Manual of mental disorders 4th Revision (DSM-IV; American Psychiatric Association, 1994), rather than as outcome measures for research purposes. The SOS (Nicholls et al., 2007) contains a category for measuring the occurrence of substance abuse with the following severity levels: 0='none', 1='occasional substance use leading to mild impairments', 2='uses illegal

substances or misuses prescription medication. Intake results in moderate adverse effect for self or others', 3='frequent substance use leading to significant physical, behavioural, emotional, relationship, occupational, or educational impairment', and 4='regular, compulsive use leading to severe impairment. Physical or psychological dependence, substance induced psychosis, mania or delirium'. However, this classification represents global impairment, rather than specific incidents of substance abuse and is thus inconsistent with the other SOS categories. Therefore, in study 6 incidents were simply coded in terms of whether substance abuse (alcohol or other illicit substance) was present based on a positive breathalyser or drug screening test, or where staff expressed suspicion of current substance abuse. Whilst this method was not validated, inter-rater reliability was excellent (K=.89).

Table 1: Measurements of adverse outcomes used in this thesis.

PAPER	OUTCOME	MEASURES
3	Aggression	Overt Aggression Scale (OAS) / Modified OAS (MOAS)
4	Aggression	Overt Aggression Scale (OAS) / Modified OAS (MOAS)
5	Aggression	START Outcome Scale (SOS)
	Self-harm/suicide	START Outcome Scale (SOS)
	Self-neglect	START Outcome Scale (SOS)
	Victimisation	START Outcome Scale (SOS)
6	Unauthorised leave	START Outcome Scale (SOS)
	Substance abuse	No standardised measure
7	Aggression	START Outcome Scale (SOS)
	Self-harm/suicide	START Outcome Scale (SOS)
8	Aggression	Overt Aggression Scale (OAS) / Modified OAS (MOAS)
	Self-harm	Overt Aggression Scale (OAS) / Modified OAS (MOAS)

2.3 Approaches to risk assessment

2.3.1 Unstructured clinical judgment

Historically, judgments concerning the future risk of engaging in adverse outcomes were determined by unstructured clinical opinion (Doyle & Dolan, 2002). This approach gave clinicians the discretion to consider what information they used to influence their decision without constraints. While this approach has advantages, including flexibility and an idiographic or person centred focus, allowing clinicians to consider unique risk factors for the individual in question (Hart, 2001), it has been criticised for being subjective and impressionistic (Grove & Meehl, 1996), having low inter-rater reliability (Hart, 1998b), showing little improvement beyond chance levels of prediction (Hart, 1998b), and leading to dispositional statements of risk, as opposed to specific statements of what may occur in a given context (Hart, 2001). Judgements made solely on unstructured clinical opinion are also subject to biases, such as being highly influenced by recent events, vivid presentation of information and personal experience, or believing that what is expected, or wanted, is more likely to occur (Carroll, 2009).

2.3.2 The actuarial approach

The contrasting approach to unstructured clinical judgment is that of actuarial risk assessment. The sole purpose of such assessments is to predict the occurrence or absence of a specific outcome, in a specific population, over a given time period. The items comprising such assessments are selected on an empirical basis, such that they have demonstrated an association with the specified outcome, and are weighted according to the strength of this association (Grann & Langstrom, 2007). This then produces a decision regarding the likelihood of the outcome occurring (Hart, 2001). Advantages to such an approach include transparency (Hart, 2001), empirical support (Hart, 2001), and inter-rater reliability (e.g., Rettenberger, Matthes, Boer, & Eher, 2010). There also appears to be little debate that they are statistically superior, in that they achieve greater predictive accuracy (Doyle & Dolan, 2002). However, they have been criticised for focusing primarily on static risk factors that offer little opportunity for intervention or treatment, and may only be useful

in informing longer term predictions of risk (Chu, Thomas, Ogloff, & Daffern, 2013) such as in informing decisions regarding discharge from secure services into the community (Bjorkly, Hartvig, Heggen, Brauer, & Moger, 2009). Further criticisms of this approach include not allowing for the consideration of idiosyncratic risk factors, exclusion of risk factors that have a logical and theoretical link with violence because they have not been proven empirically (e.g., homicidal ideation), including risk factors, such as race, which may be considered ethically and legally unacceptable, and for having "high-fidelity", such that the risk assessment tool may not perform optimally in a different assessment context (Doyle & Dolan, 2002; Hart, 1998b).

2.3.3 Structured professional judgment

The structured professional judgment approach bridges the gap between the actuarial and clinical approaches and has become the gold standard risk assessment technique (National Institute for Mental Health in England, 2004). This approach consists of evidence based guidelines that attempt to define the risk being considered, discuss qualifications for conducting the assessment, recommend what information should be considered as part of the assessment and identify a set of risk factors that should be considered (Hart, 2001). However, this approach retains flexibility in that clinicians can consider additional case-specific factors and the context in which the risk assessment is conducted (Doyle & Dolan, 2002), and they do not prescribe how the final decision is reached (Hart, 1998b). Further, there has been an increased emphasis on dynamic factors which, due to their fluctuating nature, may be better suited to short to medium term predictions (Chu et al., 2013), such as informing risk decisions whilst in inpatient treatment (Bjorkly et al., 2009). Other advantages involve improving the transparency of decision making and a shift in emphasis from risk prediction to risk management (Doyle & Dolan, 2002; Hart, 2001). The use of a structured risk assessment is now mandated in the UK (National Institute for Health and Clinical Excellence, 2005). However, guidelines specify that "no gold standard tools can be recommended" (ibid. p.18). As this thesis intends to assess whether the HCR-20 and the START are efficacious among the full range of mental health inpatients coming into contact with secure services for a wide range of adverse outcomes, it will help to

determine whether these tools are an optimal approach to risk assessment and can be recommended as the gold standard.

2.3.3.1 The Historical Clinical Risk-Management 20 (HCR-20; Webster et al., 1997; Webster et al., 1995)

A recent survey of medium secure units (Khiroya, Weaver, & Maden, 2009) revealed that the most commonly used structured professional judgment tool in the UK is the HCR-20 (Douglas et al., 2013; Webster et al., 1997; Webster et al., 1995), which was used in 79% of units. The HCR-20 is intended to aid assessments of risk for, and management of, physical violence against others in men and women over the age of 18, within correctional, civil psychiatric and forensic psychiatric settings. Violence is defined in the HCR-20 as "actual, attempted, or threatened harm to a person or persons" (Webster et al., 1997; p. 24); this definition does not take into account the severity of harm and is therefore perhaps more consistent with definitions of aggression, rather than violence, provided in section 2.2.

The tool comprises 20 items (see 5.4); ten historical items that are thought to be relatively static, five clinical items that concern current mental health related functioning, and five risk-management items that reflect clinicians' opinions of the individual's ability to adjust to the institution or community. Each item is scored on a scale of 0/No (not present), 1/Possible (possibly present or present to some degree), and 2/Yes (definitely present). Raters are then required to make a summary judgment about the likelihood (low, moderate, or high) of violence occurring based on consideration of the 20 items, and any additional case-specific factors. Applying the criteria outlined by Kraemer et al. (1997), theoretically, the historical subscale score is a multivariate fixed marker for violence, and the clinical and risk-management scales are multivariate causal factors for the same outcome. However, this classification depends on demonstration that: 1) the factors are empirically linked with violence, 2) the historical factors are indeed fixed, 3) the clinical and risk-management factors can be manipulated, and 4) the manipulations are associated with changes in the likelihood of violence occurring.

Despite its warm reception and popularity in clinical practice (e.g., Khiroya et al., 2009), there are a number of criticisms of the HCR-20, and the wider structured professional judgment approach. Such tools aim to inform management, rather than just aiding in prediction of adverse outcomes; however, with the exception of perhaps the HCR-20 companion guide (Douglas, Webster, Hart, Eaves, & Ogloff, 2001), few practical strategies for translating identified risk factors into treatment strategies have been suggested. Further, the HCR-20 and other structured professional judgment schemes tend to focus almost exclusively on the prediction of violent outcomes for which there is an extensive literature regarding risk factors (e.g., Monahan, 2002; Witt, van Dorn, & Fazel, 2013). However, in order to adequately care for and manage risks presented by mental health patients, mental health professionals must consider a broader range of clinical issues, including risks to the self, such as self-harm, suicide and self-neglect, and risk from others, in terms of victimisation (Webster et al., 2009). Whilst there are a number of outcome-specific tools available, such as the Suicide Risk Assessment and Management Manual (Bouch & Marshall, 2003) and the Self-Neglect Severity Scale (Dyer et al., 2006), it is not practical to suggest that clinicians complete a separate risk assessment tool for every possible risk faced by their patients. A final criticism pertains to the tendency of structured professional judgment tools to be dominated by risk factors at the expense of protective factors. In theory, the consideration of protective factors creates a more balanced assessment (de Vries Robbe, de Vogel, & Douglas, 2013), reducing the likelihood of negative bias, which can result in unnecessary detention (de Ruiter & Nicholls, 2011). Their consideration may also contribute towards a focus on risk-management, rather than prediction, by offering potential areas for treatment based on bolstering strengths (Nonstad et al., 2010), and may have clinical advantages in terms of promoting a therapeutic relationship (de Ruiter & Nicholls, 2011). However, whether the consideration of protective factors improves risk assessment and management or therapeutic relationships remains relatively untested.

2.3.3.1.1 Previous research with the HCR-20

The HCR-20 has consistently demonstrated its predictive ability for inpatient aggression among predominantly male samples (e.g., Arbach-Lucioni, Andrés-

Pueyo, Pomarol-Clotet, & Gomar-Soñes, 2011; Grevatt, Thomas-Peter, & Hughes, 2004; Langton, Hogue, Daffern, Mannion, & Howells, 2009; McKenzie & Curr, 2005; Morrissey et al., 2007), and meta-analyses have produced favourable results for the HCR-20 in comparison with other tools (e.g., Yang et al., 2010). However, prior to the research presented in this thesis, little was known about how the predictive efficacy of the HCR-20 may be moderated by important clinical and demographic factors. Whilst there is a wider literature pertaining to the performance of the HCR-20 in predicting outcomes post-discharge from correctional or psychiatric settings (Douglas, Shaffer, et al., 2014), these will not be reviewed due to differences in risk factors for inpatient and community outcomes (e.g., Steinert, 2002).

Conflicting results have been reported regarding the effect of gender on the predictive accuracy of the HCR-20. Nicholls et al. (2004) found that the HCR-20 historical and clinical scales predicted inpatient violence in women, but were not significant in men. In contrast, de Vogel and de Ruiter (2005) found that the HCR-20 was only predictive of violence in women when the summary judgment was used, but all of the HCR-20 scale scores, in addition to the summary judgment, were predictive in men.

Few studies have examined the role of diagnosis in the accuracy of the HCR-20 but of those that have, intellectual disability (ID) has been most commonly investigated. Fitzgerald et al. (2013) found that the HCR-20 summary judgment significantly predicted inpatient physical aggression in patients with ID but did not predict the same outcome in those without ID; none of the HCR-20 scales were significantly predictive in either group. However, effect sizes approached or exceeded the threshold for a large effect size in the ID group, with the exception of the clinical scale; therefore, the lack of predictive efficacy for scale scores in this group is probably due to small sample size (*n*=23). Morrissey et al. (2007) and Lindsay et al. (2008) found that the HCR-20 predicted physical aggression against others including inappropriate sexual behaviour, and verbal aggression/property damage in secure mental health patients with ID. However, neither of these studies compared the performance in individuals with ID to those without. With regard to other diagnoses, Tengström et al. (2006) found that the HCR-20 demonstrated greatest predictive efficacy for violence among

those with a diagnosis of schizophrenia, compared with those with cognitive impairment or personality disorder.

One study has examined the differential predictive validity of the HCR-20 as a function of ethnicity (Fujii, Tokioka, Lichton, & Hishinuma, 2005). No significant difference was found in terms of overall predictive validity between Asian Americans, Euro Americans and Native Hawaiians; however, there were significant differences in which individual items were the most potent predictors across groups. Impulsivity and young age at first violent incident were the sole predictors of aggression for Asian Americans and Euro Americans respectively. Both relationship instability and plans lacking feasibility were significant predictors for Native Hawaiians.

The studies presented above provide some evidence that the performance of the HCR-20 does differ across groups; however, none of the presented studies have controlled for additional factors that may vary between groups. Therefore, it is hard to draw firm conclusions that any observed differences in performance are due to the variable of interest. Further, there are some patient groups, such as those with organic or developmental disorders, for which the HCR-20 is untested. Given the HCR-20's popularity and the fact that it is mandated in some settings, such as St. Andrew's, its differential performance across groups warrants further attention to ensure it has predictive validity for all groups to which it is being applied; papers 1, 3, 4 and 8 address this important question.

2.3.3.2 The Short-Term Assessment of Risk and Treatability (START; Webster et al., 2009; Webster et al., 2004)

The START (Webster et al., 2009; Webster et al., 2004) is one example of recent attempts to develop risk assessment tools that include protective factors, or strengths, in response to the criticisms listed above and a strong demand from mental health practitioners (de Vries Robbe et al., 2013). The START requires the rating of 20 dynamic items on two 3-point scales (0=no/minimal, 1=moderate, and 2=high) in terms of both risk (vulnerabilities) and protective (strengths) factors (see 5.5). Raters can then identify additional case-specific factors, key strengths, critical vulnerabilities, Threats of Harm that are Real, Enactable, Acute and Targeted (THREAT) and therefore require immediate

intervention, and relevant risk history before formulating specific risk estimates for the likelihood (low, moderate, or high) of seven outcomes (violence, self-harm, suicide, self-neglect, victimisation, substance abuse, and unauthorised leave) occurring over a maximum of three months. The developers of the START do not intend the tool to replace earlier risk assessment schemes such as the HCR-20, nor is it independent from such tools; historical variables should be the foundation of any risk assessment and, therefore, the historical scale of the HCR-20 should be completed in conjunction with the START for individuals considered at risk of violence towards others (Webster, Nicholls, Martin, Desmarais, & Brink, 2006). For individuals deemed at risk in other domains, it may be necessary to consider the presence of other historical factors.

Using the criteria described by Kraemer et al. (1997), the START strength and vulnerability scales are intended as multivariate causal factors for the seven outcomes listed above; however, this claim is relatively untested. Whilst a promising development that clearly addresses a gap in available risk assessment tools, prior to the research presented in this thesis, there was no systematic review of research evaluating the START and few studies had investigated its unique contributions. Further, as the evidence base for the START is still in its infancy, empirical investigations of its performance have so far addressed its predictive ability for whole samples rather than for demographic subgroups.

Research has demonstrated that the START can predict inpatient aggression (e.g., Chu, Thomas, Ogloff, & Daffern, 2011; Desmarais, Nicholls, Wilson, & Brink, 2012; Gray et al., 2011), but prior to this thesis little was known about its ability to predict the other adverse outcomes intended by the tools' authors, or whether the inclusion of strengths has any additional benefits over that achieved by the consideration of risk factors. Whilst research has suggested that clinicians' perceive individuals to be at risk in multiple domains, which can be evidenced by significant correlations between specific risk estimates for violence and all other outcomes (Nicholls, Petersen, Brink, & Webster, 2011), it is unclear to what extent these outcomes are related, or share common risk factors. The START authors cite reasons for including different factors in the START manual (Webster et al., 2009), but rarely provide empirically supported

28

links between factors and adverse outcomes beyond violence. Therefore, it is important to assess the predictive validity of the START for the full range of outcomes it intends to predict, in order to assess if its scales are a multivariate causal factor for all these outcomes and to consequently determine the feasibility of using one risk assessment tool to assess multiple outcomes. Papers 2, 5, and 6 address the above question; in addition, paper 7 is the first investigation of whether gender affects the predictive accuracy of this tool.

2.4 Data collection

When conducting risk assessment research, researchers must decide whether to use routinely collected risk assessment and outcomes data, or have members of the research team collect such data; both approaches have advantages and disadvantages. A further consideration involves whether to collect data prospectively or use retrospective data; in order to enable judgements regarding predictive accuracy it is necessary to approximate a prospective design when using retrospectively collected data (i.e., a pseudo-prospective design) such that risk assessments are completed based only on information available prior to the period over which outcome data is obtained. Advantages of using routinely collected data include access to much larger datasets, in a way that is relatively low-cost and resource light, as the cost of using such data may be limited to that involved in analysis and dissemination (Bain, Chalmers, & Brewster, 1997). Also, routinely collected data often covers a whole population, as opposed to a sample (Bain et al., 1997) and therefore distribution of demographic characteristics, incidence, and prevalence are truly representative of these values for the population in question. Finally, using risk assessments that are routinely collected as part of clinical practice has greater ecological validity than using assessments completed by researchers and therefore gives a more accurate representation of the likely efficacy of risk assessment schemes when used as intended.

Routinely collected data was used for all primary studies presented in this thesis (papers 3-8) for the reasons identified above. However, it may be difficult to establish the quality of the routinely collected data, such as its completeness (i.e., the extent to which every adverse incident is captured; Bain et al., 1997),

and the inter-rater reliability of coded risk assessment and outcomes data. In instances where the reliability of a measure has already been established, as is the case with the HCR-20 and the START (e.g., Desmarais, Collins, Nicholls, & Brink, 2011; Douglas, Shaffer, et al., 2014), the reliability of the risk assessments may be less of a concern. However, appropriate training should be provided for all individuals conducting risk assessments to ensure that a standardised approach is used, as was the case in the current study setting (see 5.11 and 5.13 for a description of START training provided). Further, for all included papers the outcomes data was coded by members of the research team using standardised measures, after inter-rater reliability had been established. Concerns regarding the completeness of the data may be alleviated by the fact that higher rates of aggression were observed in the papers presented in this thesis (e.g., O'Shea, Picchioni, & Dickens, 2016; O'Shea, Picchioni, Mason, Sugarman, & Dickens, 2014a), compared with previous research. However, this was not the case for the remaining outcomes. It is unclear whether this is due to real differences in rates of adverse outcomes across studies, or a greater propensity for recording aggressive, as opposed to other, adverse outcomes in the setting where data for the presented papers was collected. A final issue with the reliance of routinely collected data is that insufficient information regarding methods may be available to allow for replication (de Lusignan & van Weel, 2006).

A further consideration when collecting outcome data is selecting an appropriate length for follow-up. All of the original empirical papers presented in this thesis (i.e., papers 3-8) examined outcomes of interest over a three-month follow-up period. This was selected to balance out the need to allow sufficient time for the occurrence and detection of events of interest with the resources required to code adverse incidents. Also, a shorter follow-up period is warranted in the inpatient setting as assessments are conducted more regularly to reflect changes in risk over the course of treatment. It is possible that longer follow-up periods would have resulted in greater predictive validity as there is greater opportunity for events of interest to occur and more opportunity for events to be detected (Leistico, Salekin, DeCoster, & Rogers, 2008). Many risk assessment schemes come with recommended follow-up times; for example, the START

manual recommends that the assessment is repeated at least every three months (Webster et al., 2009). However, researchers may wish to examine various follow-up periods unless it has been established empirically that the reported follow-up time is optimal. Examination of the optimal follow-up period for the START has revealed that the 3-month recommendation is appropriate; however, some individuals (e.g., those rated at elevated risk who have not engaged in any relevant risk behaviours over a 2-month period) may benefit from earlier re-assessment (Dickens & O'Shea, 2015).

2.5 Analysis

Methods of evaluating predictive or diagnostic test accuracy typically assume that a test, or risk assessment, has a positive result (indicating that a diseased state exists or an adverse outcome is likely to occur), or a negative result (indicating that a diseased state does not exist or that an adverse outcome will not occur) (Mossman, 2013). Also, it is assumed that the outcome is dichotomous (Mossman, 2013). Therefore, indicators of predictive accuracy are typically calculated from a 2x2 contingency table (Singh, 2013), which represent assessment and outcome information in the form of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) (see Figure 2). Singh (2013) distinguishes between measures of global accuracy, such as Area Under the Curve (AUC) values and logistic odds ratios, and measures of high risk accuracy (the ability to identify high risk groups, specifically) and low risk accuracy (the ability to identify low risk groups).

<table>
<tr><td></td><td></td><td colspan="2">Outcome</td></tr>
<tr><td></td><td></td><td>+</td><td>-</td></tr>
<tr><td rowspan="2">Risk Assessment</td><td>+</td><td>TP</td><td>FP</td></tr>
<tr><td>-</td><td>TN</td><td>FN</td></tr>
</table>

Figure 2: 2x2 contingency table

2.5.1 Sensitivity and Specificity

Sensitivity and specificity can be calculated from information in the 2x2 contingency table presented above (see Table 2 for formulas). Sensitivity is a

measure of the ability to discriminate high risk individuals and is the proportion of positive outcomes that are correctly identified by the assessment (Altman & Bland, 1994a; Singh, 2013). Specificity is a low risk discrimination index that represents the proportion of negative outcomes correctly identified (Altman & Bland, 1994a; Singh, 2013). There are two key limitations to sensitivity and specificity; first, that they are affected by the base rate of the behaviour or disorder of interest such that sensitivity tends to increase, and specificity decrease, as the base rate increases (Singh, 2013). Second, sensitivity and specificity are intended for use when a measure has a single cut-off point, such that scores above this threshold indicate a positive test result, which is rarely the case with risk assessment tools where individuals are classified into multiple risk categories based on item scores and clinical judgement (Singh, 2013). For this reason, sensitivity and specificity are not suitable for assessing the accuracy of the HCR-20 and the START as they are not intended to have cut-off points, such that clinicians can deem an individual to be high risk in the absence of a large number of risk factors; therefore, they were not reported in the papers included in this thesis. However, they are of relevance as they form the basis for the calculation of AUC values, which are the primary outcomes for papers 3-8.

Table 2: Formulas for the calculation of indicators of predictive accuracy

MEASURE OF PREDICTIVE ACCURACY	FORMULA
SENSITIVITY	$\dfrac{TP}{TP + FN}$
SPECIFICITY	$\dfrac{TN}{TN + FP}$
POSITIVE PREDICTIVE VALUE	$\dfrac{TP}{TP + FP}$
NEGATIVE PREDICTIVE VALUE	$\dfrac{TN}{TN + FN}$
NUMBER NEEDED TO DETAIN	$\dfrac{1}{PPV}$
NUMBER SAFELY DISCHARGED	$\left(\dfrac{1}{1 - NPV}\right) - 1$

2.5.2 Positive and Negative Predictive Values

Positive Predictive Values (PPVs) refer to the proportion of those classified as test outcome positive, or high risk, that go on to engage in the behaviour of interest whereas the Negative Predictive Value (NPV) is the proportion of those judged as low risk, or test outcome negative, that do not engage in the behaviour of interest (Singh, 2013), and can be calculated from the formulas presented in Table 2. Like sensitivity and specificity, PPV and NPV are limited by reliance on a single cut-off point (Singh, 2013). Further, they are strongly affected by base rates such that PPVs will not approach 1 if the base rate of the behaviour is very low, even if sensitivity and specificity are high (Altman & Bland, 1994b). Therefore, it is not possible to compare PPV and NPV values across studies except in instances where base rates are consistent across populations (Singh, 2013). PPVs and NPVs were not reported as primary outcomes in the included papers due to the above limitations. However, they can provide useful information regarding whether a tool is more accurate at identifying high or low risk individuals. Therefore, PPVs and NPVs of the specific risk estimates were reported in papers 5 and 6.

2.5.3 Receiver Operating Characteristics and Area Under the Curve values

Receiver Operating Characteristic (ROC) analysis emerged in the 1950s in the context of signal detection and problems with radar; in the 1960s, it started being used in medical imaging (Zweig & Campbell, 1993). The first article to apply ROC analysis to violence prediction (Mossman, 1994) was not until the mid-1990s; however, by the mid-2000s, ROC analysis had become the standard technique for evaluating tools aimed at assessing risk of violence (Mossman, 2013). Prior to this, studies of violence prediction typically reported sensitivity, specificity, PPV, and NPV as indicators of predictive accuracy.

ROC curves (see Figure 3) depict the accuracy of a tool or judgment by plotting the true positive rate, or sensitivity, against the false positive rate (1-specificity) for the complete range of decision thresholds, or cut-off points (Mossman, 2013); a curve that passes through the top left hand corner of the graph would have perfect sensitivity and specificity. Area Under the Curve (AUC) values are

the most commonly used summary index of ROC analysis; it can be interpreted as the probability that a randomly selected violent individual will have been assigned a higher probability of violence, or have a higher score on a risk assessment measure, than a randomly selected non-violent individual. The AUC parameter ranges from 0.00 (the non-violent individual always has a higher score than the violent individual) to 1.00 (perfect discrimination: the violent individual always has a higher score than the non-violent individual). An AUC value of 0.50 is equivalent to chance prediction (i.e., the violent individual has a higher score than the non-violent individual 50% of the time; therefore, scores do not discriminate between groups and predictive value is equivalent to tossing a coin). There is some debate in the literature about what constitutes a small, moderate, or large effect size. Dolan and Doyle (2000) state that AUC values greater than 0.75 (i.e., the violent individual has a higher score than the non-violent individuals 75% of the time) are considered large, whereas Douglas and Reeves (2010) suggest that those over 0.70 may be considered large. AUC values of 0.56, 0.64, and 0.71 correspond to small (0.20), moderate (0.50), and large (0.80) Cohen's (1992) *d* values, respectively (Rice & Harris, 2005), which arguably are the most commonly used measure of effect size (Kraemer & Kupfer, 2006; Rice & Harris, 2005). The papers presented in this thesis classified effect sizes as large according to the more conservative value of 0.75 suggested by Dolan and Doyle (2000).
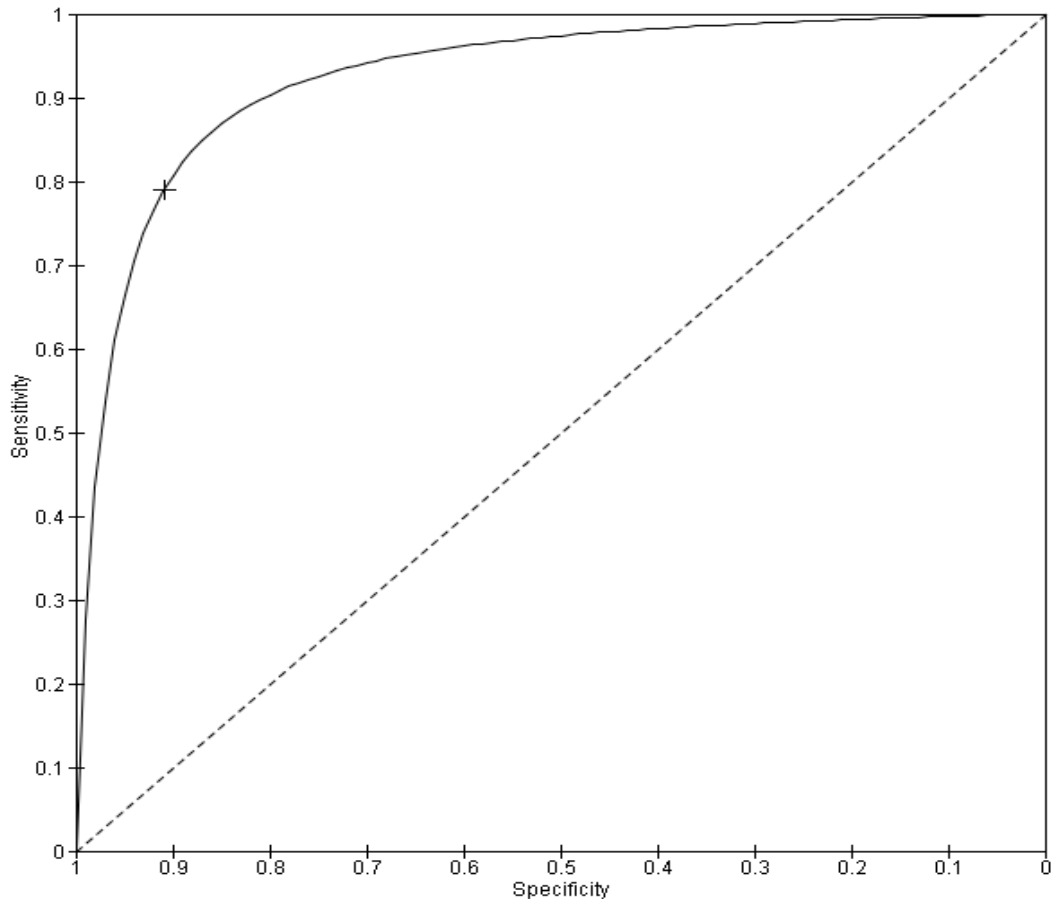
Figure 3: Example Receiver Operating Characteristic (ROC) curve

There are a number of advantages to the ROC approach which have contributed to it becoming the method of choice for assessing predictive efficacy. As ROC analysis is based on true positive rates and false positive rates, which are not related to the base rate of the outcome in question, ROC analysis is similarly independent of base rates (Mossman, 2013). When behaviours have a low base rate (i.e., occur infrequently), they are inherently difficult to predict. Measures of accuracy such as positive predictive values will likely be low as the proportion of those scoring as test positive (i.e., high risk) will be dominated by the false-positive group (Szmukler & Rose, 2013). However, it would be inappropriate to ignore information regarding base rates of behaviour when making practical decisions regarding interventions, detention, and resource allocation (Singh, 2013). Further, ROC methods are unaffected by clinicians' biases regarding Type I or Type II errors (Mossman, 1994), and scoring methods of the risk assessment (Pepe, Janes, Longton, Leisenring, &

35

Newcomb, 2004), aiding comparisons between AUC values derived from different tools or methods.

However, summarising sensitivity and specificity across multiple cut-off points into a single AUC value provides no information regarding the balance between false negative and false positive errors (Mossman, 2013). Also, AUC values have been found to be smaller when there is a restricted range of scores (Hanson, 2008); this would be expected in the inpatient settings as samples are pre-selected as high risk (i.e., individuals would not be detained if they were not deemed a risk to themselves or others). Further, AUC values may be best considered as a measure of discriminatory ability, rather than predictive accuracy; if all individuals that engaged in the behaviour of interest scored higher than all individuals that did not engage in the behaviour, this would produce a perfect AUC value of 1, even if none of the individuals scored particularly highly, or were judged as high risk (Singh, 2013).

Due to the advantages presented above, and to facilitate comparisons with previous research into predictive efficacy of assessment tools, AUC values were the primary outcome for papers 3-8. However, for papers 3, 4, 7 and 8 the AUC values were derived from rocreg analysis. Rocreg analysis extends traditional ROC analysis by allowing for the incorporation of covariates, which may impact the discriminatory accuracy of risk assessment tools such that they affect the separation between scores of violent and non-violent individuals (Janes, Longton, & Pepe, 2009), and the modelling of covariate-specific ROC curves. Further, rocreg analysis is able to control for the effect of additional covariates that may alter the discriminatory accuracy and/or score distribution of risk assessment schemes (Janes et al., 2009). This method therefore allowed for the examination of whether the accuracy of the HCR-20 and the START varied as a function of patient characteristics (e.g., gender) while controlling for other relevant characteristics that differed between groups (e.g., diagnosis), based on inspection of covariate-specific ROC curves and their corresponding AUC values. Despite their promise, these methods have scarcely been used, perhaps due to the methods being relatively recently developed, and the lack of available ROC regression software (Rodríguez-Álvarez, Tahoces, & Cadarso-Suárez, 2011). In recent years, these methods have gained use in the medical

literature (e.g., Lerner et al., 2015; Mathé et al., 2014; Motamed et al., 2015; Walsh et al., 2012). However, the current papers presented in this thesis represent the first time such methods have been used in the risk assessment of behavioural outcomes. This extends previous risk assessment research by increasing the confidence that can be placed in the results that any observed differences in performance between groups are actually due to the variable of interest.

2.5.4 Odds ratios

Odds ratios represent the increase in the odds of an outcome of interest occurring for those rated at increased risk compared to those rated low risk, or for every one-point increase on a continuous scale, such as a score on a risk assessment measure. Calculation of logistic odds ratios also allows for the examination of how other variables affect the odds ratio using logistic regression (Bland & Altman, 2000; Singh, 2013). Odds ratios of 1.50, 2.50, 4.00, and 10.00 can be considered small, moderate, large, and very large respectively (Rosenthal, 1996). Like AUC values, odds ratios are unaffected by changes in the base rate of the outcome of interest (Singh, 2013). However, unlike AUC values they are unaffected by a restricted range of scores so may be more accurate in samples that are considered high risk (Hanson, 2008). Further, it has been suggested that odds ratio may be more familiar to clinicians, and therefore easier to understand, compared to other indicators of predictive accuracy (Singh, 2013). For these reasons, odds ratios were calculated to supplement presented AUC values for papers 5, 6, and 7 of this thesis.

2.5.5 Number Needed to Detain and Number Safely Discharged

Number needed to detain (NND) and Number Safely Discharged (NSD) are recently developed indicators of predictive accuracy and are calculated from PPVs and NPVs. The NND can be defined as the number of individuals judged as high risk or test outcome positive that would require detaining in order to prevent a single incident of the behaviour of interest occurring in the community (DeClue & Campbell, 2013; Fazel, Singh, Doll, & Grann, 2012). However, in the inpatient setting the NND could be considered as the number of high risk

individuals who would require risk management strategies targeting a specific behaviour to prevent one incident occurring. Similarly, the NSD is typically defined as the number of low risk individuals that could be discharged prior to a single incident occurring in the community (DeClue & Campbell, 2013; Fazel et al., 2012), but in the inpatient setting should be considered as the number of low risk individuals that could safely be managed without risk management strategies targeting the behaviour of interest prior to an incident occurring. NNDs and NSDs are limited in the same ways as sensitivity, specificity, PPVs, and NPVs, such that they are affected by base rates of the behaviour of interest and are designed for use with a single cut-off point. Further, Singh (2013) states that they may be difficult to interpret as they involve an ethical judgment that requires balancing public protection against civil rights; imposing restrictions for five people to prevent one violent incident may be considered acceptable to some, whereas others may feel this is an unnecessary restriction of civil rights. NNDs and NSDs were calculated for paper 6 of this thesis to supplement the AUC values and provide information to facilitate risk management decisions.

2.5.6 Conclusion

The statistical methods presented above have both strengths and limitations; whilst AUC values are the preferred method for evaluating predictive accuracy (Mossman, 2013), they do not give a complete picture of a tool's performance as they are a global measure of discrimination, rather than of the tool's calibration (Singh, 2013). Supplementing the AUC with the additional measures presented above may provide a more complete picture of the performance of risk assessment tools by providing information about high and low risk discrimination; this was done for papers 5, 6 and 8 presented in this thesis. However, this required the combination of different levels of risk to establish a single cut-off point (Singh, 2013), which may be an oversimplification of the included risk assessment tools. Singh (2013) suggests that the development of a set of calibration indicators (i.e., measures that can be used to evaluate the accuracy of a tool) that are independent of cut-off points would be beneficial for the future of risk assessment research; such a development would allow for a more fine-grained analysis of the accuracy of the HCR-20 and the START than was possible in the current papers.

2.6 Reporting guidelines

Translating research findings into tangible benefits and recommendations for future research depends on the dissemination and publication of high quality, reliable data (Simera, Moher, Hoey, Schulz, & Altman, 2010). Reporting guidelines and checklists, which provide a set of rules or principles to promote accurate reporting, have been developed to combat growing evidence of deficient reporting (Simera et al., 2010).

2.6.1 The Standards for Reporting of Diagnostic Accuracy (STARD; Bossuyt et al., 2003)

The STARD is a 25 item checklist to improve the accuracy and completeness of reporting of studies of diagnostic accuracy. The items were derived from a systematic review of publications on the conduct and reporting of diagnostic studies and narrowed down to a list of 25 items during a consensus meeting of the STARD steering group (Bossuyt et al., 2003). The STARD checklist contains items relating to all aspects of the manuscript (title/abstract/keywords, introduction, methods, results, discussion) and addressed areas such as recruitment and data collection, rationale and conduct of the reference standard, methods of analysis, and comprehensive reporting of results (Bossuyt et al., 2003). The STARD has not routinely been used in risk assessment research, due to its primary focus on diagnostics; however, it is an important precursor to The Risk Assessment Guidelines for the Evaluation of Efficacy (Singh, Yang, Mulvey, & RAGEE Group, 2015) described below.

2.6.2 The Risk Assessment Guidelines for the Evaluation of Efficacy (Singh et al., 2015)

The RAGEE checklist was developed in response to evidence that research examining the predictive validity of risk assessment tools is not as transparent and accurately reported as research areas with established reporting guidelines (Singh et al., 2015). Further, it aimed to address the limitations of reporting guidelines for diagnostic accuracy, such as the STARD, when applied to risk assessment research. The RAGEE items were developed through systematic searches to identify existing reporting guidance for prognostic and diagnostic accuracy studies and Delphi techniques with 37 experts in violence risk

assessment (Singh et al., 2015). The resulting statement comprised 50 items indicating the minimal information that should be reported for risk assessment studies: 4 items for the abstract, 2 for the introduction, 30 for the methods (split into items regarding participants, instrument design, instrument administration, study design, predicted outcome, and statistical analysis), 6 items for results (split into participant outcomes and predictive validity), 4 items for the discussion, and 4 items related to recommended disclosures (Singh et al., 2015). Initial investigations indicate experts in the area were satisfied with the checklist and supported its use as reporting standards for research into the predictive validity of violence risk assessments (Singh et al., 2015). A number of the papers presented in this thesis precede the availability of the RAGEE guidelines; however, these principles were followed where possible. Adherence to these guidelines in future research would improve quality and facilitate comparisons across papers.

2.7 Ethical considerations

When conducting risk assessment research there are a number of considerations to be made regarding ethics and confidentiality. First, appropriate ethical approval must be obtained. Research that involves direct contact with the individuals who are the subject of the research must include individual informed consent, as must research that involves the release of personally identifiable data from existing databases (Bain et al., 1997). Approaching individual patients and obtaining informed consent to access their records for the purpose of the research presented in this thesis was considered. However, as included individuals would then be self-selecting, this would have likely resulted in a much smaller and less diverse sample than that obtained through the use of routinely collected data. This would have been damaging to the quality of the presented research and may have limited the ability to investigate differential predictive validity.

In instances where researchers use routinely collected data that is provided in an anonymised form, ethical approval is not typically required (Bain et al., 1997; DH Research and Development Directorate [England], National Institute for Social Care and Health Research [Wales], Chief Scientist Office [Scotland], &

R&D Division Public Health Agency [Northern Ireland], 2011). Therefore, the decision was made to use HCR-20 risk assessment data that had been collected in an anonymised form for the purpose of an audit of HCR-20 completion across St. Andrew's, for which the candidate was hired to conduct; this would provide a large sample that was truly representative of the population. In order to collect this data, the candidate designed a template for data collection using Microsoft Excel (see 5.6) and populated this with the relevant patient details for each responsible clinician, which had been provided to the candidate in an anonymised form. These spreadsheets were then circulated to the responsible clinicians (with one sheet relating to each patient under their care) and they were asked to fill out the relevant HCR-20 information. Some clinicians were able to complete these forms and return them to the candidate within the required timeframe; however, there was a large amount of data outstanding. To collect the remaining data and minimise the demands placed on clinicians, the candidate arranged a time to meet with clinicians who had not returned data for their patients who then read out scores for the HCR-20 items so that these could be recorded by the candidate. As progress notes (required for coding outcomes data) and START risk assessments were entered in an electronic format in the study setting, the IT department was able to provide the candidate with this information in an anonymised format via the creation of an automated report; outcome and risk assessment data was then linked by the candidate using the unique patient numbers. Therefore, as all of the data seen by the researchers was anonymised, and no new data was collected for the purpose of the research projects that form the basis of this thesis, formal ethical approval from an NHS Research Ethics Committee was not obtained. Further, as the data was collected prior to the commencement of the PhD, the studies were not subject to the university's ethics committee. However, the projects were supported by the Chief Executive Officer and Medical Director at St. Andrew's Healthcare and formally approved by the Head of Clinical Effectiveness.

The imperfection of predictions regarding future adverse events raises questions as to the decisions that can be reasonably informed by the results of risk assessments. Ethical concerns are particularly prevalent in situations where

there are serious consequences to being deemed as high risk, such as in legal contexts or when making admission and discharge decisions regarding secure care, as it increases the likelihood that individuals who would have not engaged in future adverse events will be included among those facing restrictions and long-term loss of liberty (Prentky, Janus, Barbaree, Schwartz, & Kafka, 2006). It has therefore been suggested that their use should be restricted to situations where the consequences of errors are minimal, such as priority for admission into a treatment program; in contrast, it is unlikely that they meet the legal criteria for admissibility of expert evidence in court proceedings (Hart, Michie, & Cooke, 2007; Prentky et al., 2006). As all the risk assessments used in this research were conducted as part of routine clinical practice, as opposed to for the studies themselves, the conduct of this research did not have a direct impact on individuals. However, this thesis will help shed light on the outcomes for which risk assessment tools have sufficient predictive validity and are therefore likely to be informative, and those for which their efficacy is limited and it may be inappropriate to base decisions on their results. Therefore, the results of this research may affect future clinical practice both in the current study setting and more widely.

Ethical issues also impose restrictions on the type of study designs permissible. Randomised controlled trials (RCTs) are considered the optimal study design for assessing the effectiveness of interventions (Ryan et al., 2013). A small number of RCTs have been conducted in this area (e.g., Abderhalden et al., 2008; Troquete et al., 2013), but it may be considered unethical not to assess risk in secure mental health patients and the use of risk assessment schemes is considered best practice in secure mental health care in the UK (National Institute for Mental Health in England, 2004). Many individual organisations also mandate their use, rendering RCTs of this type extremely difficult. Hart (1998b) argued that once a risk has been identified clinicians are obligated to do everything in their power to prevent the occurrence of negative events; however, this article was concerned with violence, which was defined earlier in this thesis as aggression that is intended to cause severe harm (Anderson & Bushman, 2002). Where other risks are considered, it may involve balancing available resources and avoiding unnecessary restrictions with the nature and

severity of the adverse event. Either way, interventions by clinical staff may create the impression that an incorrect level of risk was assigned to the individual. It is, therefore, difficult to accurately establish whether risk assessments and the management strategies that result are effectively reducing the occurrence of negative events.

There are also ethical issues concerning the selection of risk factors to be included in risk assessment schemes. Both gender and race have been shown to be correlated with adverse outcomes such as violence; however, many find their inclusion morally objectionable as it will inevitably contribute to increased discrimination (Gottfredson & Moriarty, 2006). Further, risk assessments based on such factors are of limited use as one-off passive predictions, as clinicians cannot target the fact that someone is, for example, male with risk management strategies (Hart, 1998b). However, omitting such factors, as is common practice, has also been criticised due to the resulting loss of predictive accuracy. Unfortunately, a suitable alternative strategy has not been reached (Gottfredson & Moriarty, 2006).

2.8 Summary

This chapter has provided context for the empirical papers that are presented in the following chapter by presenting background information regarding methodology for risk assessment research in secure mental health settings. While not intended to be an exhaustive review of available methods and measures, the chapter highlights the complexities involved in research in this area and provides rationale for the methodological decisions made in the presented papers. In detail, the above sections have: 1) provided an overview of the importance of accurately defining risk factors, protective factors, and adverse outcomes, 2) outlined a number of approaches to measuring such outcomes and provided the rationale for the measures used in the presented papers, 3) provided some background regarding different approaches to risk assessment and previous research with the HCR-20 and the START, which are the focus of this thesis, 4) reviewed advantages and disadvantages of methods for data collection, 5) discussed current statistical methods and how these could be improved, 6) summarised the importance of reporting guidelines for

improving quality and consistency of research, and 7) outlined necessary ethical considerations. A secondary aim of this chapter is to provoke consideration of these issues by researchers with the hope of improving the standard of future investigations in this area.

3   Summary of included papers

3.1   Paper 1: Moderators of the predictive efficacy of the HCR-20: Systematic review and meta-analysis (O'Shea, Mitchell, Picchioni, & Dickens, 2013)

This paper presents a systematic review and meta-analysis of the existing literature pertaining to the predictive ability of the HCR-20 (version 2) for inpatient aggression and self-harm within a secure mental health setting. This work was conducted in order to address the first part of research question 1 (see Figure 1), i.e., whether the predictive efficacy of the HCR-20 varies based on demographic and clinical characteristics (see 5.7 for full paper). This was a necessary step as no previous reviews of the HCR-20 have examined moderators of its predictive validity and there has been a lack of primary research regarding differential predictive validity across groups; further, research regarding the influence of gender has been conflicting (see section 2.3.3.1.1). It has a secondary aim of examining whether performance differs based on which subscale is used, the type of aggressive outcome being predicted, and methodological factors such as length of follow-up.

Comprehensive search terms were used to search seven electronic databases for articles published between January 1995 and August 2012; additional papers were located through handsearching. This strategy resulted in 20 non-overlapping datasets including a total of 2,067 participants. This paper was the first of its kind to examine evidence on the HCR-20 in isolation from other risk assessment tools, allowing for the investigation of potential moderators. Gender, diagnosis, ethnicity, length of follow-up, independence of outcome assessment from risk assessment, and identified risk of bias moderated HCR-20 effect sizes. Effect sizes were larger in samples containing higher proportions of patients with schizophrenia, males, and Caucasians; effect sizes were smaller when there were higher proportions of individuals with personality disorder. Further, effect sizes were larger in studies with a higher risk of bias, longer follow-up periods, and prospective study designs. A key limitation of this paper is that it was not possible to fully investigate potential moderators due to insufficient heterogeneity across included studies, low reporting rates of clinical

and demographic characteristics of samples, and the fact that data for different groups of patients are usually aggregated.

3.1.1 Author's contribution

The International Committee of Medical Journal Editors recommends that authorship should be based on the following criteria: 1) substantial contributions to study conception and design, or acquisition, analysis, or interpretation of data; 2) drafting the work or revising it for intellectual contribution; 3) approval of the final version for publication; and 4) agreeing to be accountable for the integrity and accuracy of the work. Following these guidelines, the candidate made a substantial contribution to designing the search strategy and selecting inclusion criteria. The candidate extracted data from selected papers and assessed study quality in collaboration with other authors. The candidate was responsible for data analysis and completed the first draft of the article. The article was revised for intellectual contribution by the candidate and co-authors, and the candidate approved the final version for publication. This was deemed to represent 40% of the total work on this paper.

3.2 Paper 2: Short-Term Assessment of Risk and Treatability (START): Systematic review and meta-analysis (O'Shea & Dickens, 2014)

This paper, which aims to address the first part of research question 2 (see Figure 1), documents the first systematic review and meta-analysis of the START (see section 2.3.3.2 for an overview of the START research conducted prior to this thesis) and aimed to examine all aspects of the START, especially its relatively unique focus on outcomes other than aggression and violence and the inclusion of protective factors and whether these have incremental validity (see 5.8 for full paper). Five electronic databases were searched using comprehensive terms for articles published up until January 2013; additional papers were located through handsearching and reviews of reference lists. Twenty-three studies were eligible for inclusion in the narrative review and nine studies (N=543) were included in the meta-analysis. The START demonstrated high internal consistency, interrater reliability and convergent validity with other assessments of risk and protective factors. There was a lack of available

information regarding the ability of the START to document change, although users felt that the START was useful in this regard (Desmarais et al., 2011; Kroppan et al., 2011). Users generally felt they had sufficient time to score the START, information was readily available, and endorsed statements regarding its utility. Confidence ratings were high, although users expressed difficulty in completing specific risk estimates, risk formulations and signature risk signs.

Results of the meta-analysis and narrative review combined suggest that the START is a stronger predictor of aggressive outcomes than the remaining outcomes. There is some evidence that the specific risk estimate for self-harm can predict the corresponding outcome, but effect sizes for self-neglect and victimisation were not significantly different from chance. There was insufficient data available to calculate mean weighted effect sizes for unauthorised leave, suicide, and substance abuse. Further, insufficient evidence was available to ascertain whether the inclusion of strengths improves upon prediction achieved by the vulnerability scale. The major limitation of this paper is that it was not possible to fully investigate some of the unique factors of the START due to limited data; however, this is as much a criticism of the dearth of the literature, as it is of the current study itself. Further, it was not possible to investigate potential moderators due to the small number of studies included. However, given that paper 1 of this thesis demonstrated that the HCR-20 is affected by such factors, it would be reasonable to assume that the START may be similarly affected.

3.2.1 Author's contribution

The candidate made a substantial contribution to designing the search strategy and selecting inclusion criteria, extracting data from selected papers, and assessed the quality of included articles in collaboration with her co-author. The candidate was responsible for data analysis and completed the first draft of the article. The article was revised for intellectual contribution by the candidate and co-author, and the candidate approved the final version for publication. The candidate and her co-author agreed that they made equal contributions to this paper.

3.3 Paper 3: Differential predictive validity of the HCR-20 for inpatient aggression (O'Shea et al., 2014a)

This study was a further examination of the research question addressed by paper 1 (see section 3.1). Paper 1 revealed that clinical and demographic factors do affect performance of the HCR-20 in the inpatient setting, but was only able to examine performance based on the proportion of demographic and clinical characteristics among samples. Paper 3 (see 5.9) therefore expanded upon the results of the meta-analysis by directly examining how the performance of the HCR-20 is moderated by demographic and clinical characteristics (namely diagnosis, gender, age and ethnicity) whilst controlling for potential covariates using rocreg analysis. This paper also starts to address the second half of research question 1 (how variations in the performance of the HCR-20 impacts treatment planning; see Figure 1) through the examination of differences in the predictive validity of the individual HCR-20 items across groups.

A pseudo-prospective examination of the predictive ability of the HCR-20 for any aggression and physical aggression occurring in the three months following assessment was conducted (N=505). The HCR-20 performed more accurately in women compared with men, and in those with schizophrenia and/or personality disorder compared with those with organic or developmental disorders; age and ethnicity differences were less robust. Examination of the relevance of individual HCR-20 items to different groups suggested that risk-management items may be of most relevance to women, while clinical items are more important in men. Most of the clinical and risk-management items were predictive of aggression in the schizophrenia and personality disorder groups, but fewer items were relevant for the remaining diagnostic categories.

One of the main limitations of this paper was the relatively short follow-up period. As discussed in section 2.4, superior performance may be expected over longer follow-up periods; this was found to be the case in paper 1 (discussed in section 3.1). However, this allowed for data collection relating to a much larger sample of patients than has previously been used and therefore allowed for the direct investigation of moderating influences; the largest sample

size among the HCR-20 inpatient literature prior to this study was 238 (McDermott, Quanbeck, Busse, Yastro, & Scott, 2008).

3.3.1 Author's contribution

The candidate was responsible for designing templates for data collection and acquiring data, which was collected as part of an organisation-wide audit of HCR-20 risk assessment completion. The candidate was also responsible for analysing data and completed the first draft of the paper. The candidate contributed to revising the manuscript in collaboration with co-authors, and approved the final version for publication. This was judged to amount to 40% of the work on this paper.

3.4 Paper 4: The effect of intellectual disability on accuracy of the HCR-20 (O'Shea, Picchioni, McCarthy, Mason, & Dickens, 2015)

People with ID account for a disproportionate amount of the aggressive incidents that occur within secure settings (Dickens, Picchioni, & Long, 2013). Therefore, it is important to determine how the HCR-20 performs within this population. This study extends the work of papers 1 and 3, which showed that the efficacy of the HCR-20 differs based on sample characteristics, to examine the effect of ID on its performance and the implications this has for risk management and treatment planning. It was the first study to examine the performance of the HCR-20 in patients with ID (n=109) compared with mental health inpatients with a mental disorder, but no ID (n=504), whilst controlling for characteristics that differ between the two groups (see 5.10 for full paper). The same study design and analytic procedures were used as in paper 2.

Consistent with previous research, significantly higher proportions of those in the ID group engaged in physical aggression and any aggression compared to the comparison group. The HCR-20 predicted aggression in both groups. The clinical subscale performed significantly better in those without an ID compared to those with an ID; however, the clinicians' summary judgment was only significantly predictive for the ID population.

A key limitation of this study was the lack of available information regarding IQ; therefore, we could not examine whether predictive validity of the HCR-20

differed within the overall ID category. Patients in the current study had mild to moderate ID; hence, these findings should not be generalised to those with severe ID. Item-outcome analyses revealed that clinical items were most relevant to the comparison group and historical items were most relevant to the ID groups; it is possible that there is greater variation in historical scale scores in the ID group, allowing it to act as a meaningful predictor in this population. Interventions for reducing aggression in individuals with an ID should aim to limit exposure to destabilisers, or increase tolerance to such factors.

3.4.1   Author's contribution

As with paper 2, data was collected by the candidate as part of an audit. The candidate was responsible for analysing data and completed the first draft of the paper. The candidate contributed to revising the manuscript in collaboration with co-authors, and approved the final version for publication. As with the previous paper, this amounted to 40% of the total work.

3.5   Paper 5: The predictive validity of the Short-Term Assessment of Risk and Treatability (START) for multiple adverse outcomes in a secure psychiatric inpatient setting (O'Shea et al., 2016)

Paper 2 (see section 3.2) highlighted a lack of evidence for the START as a predictor of outcomes beyond aggression and violence. Therefore, this study aimed to expand current knowledge, and extend the investigation of research question 2 (see Figure 1), by investigating the predictive validity of the START for the full range of adverse outcomes it aims to predict. Further, it aimed to determine whether the strength scale has incremental validity over the vulnerability scale. This pseudo-prospective examination (N=200) revealed that the START was a robust predictor of aggressive outcomes, but did not appear to outperform the HCR-20, although no comparison was included. Strength scores had incremental validity over the vulnerability scores for these outcomes, suggesting that their consideration adds unique predictive ability. START strength scores significantly predicted self-neglect but neither scale predicted self-harm/suicide or victimisation. Since it is the strength rather than the vulnerability scores that predict self-neglect, interventions should target improving skills rather than reducing deficits. The lack of relevant factors for

self-harm/suicide and victimisation suggest that further development is required to better establish risk factors for diverse outcomes.

The specific risk estimates for self-harm and suicide predicted the related composite outcome and the corresponding risk estimate for victimisation predicted its occurrence. The finding of significant prediction by the specific risk estimates in the absence of significant prediction achieved by the scores, combined with evidence that the specific risk estimates have incremental validity over scale scores, suggests that clinicians are supplementing the scores with their clinical knowledge and expertise (see 5.11 for full paper).

It was not possible to examine the predictive efficacy of the START for unauthorised leave and substance abuse due to low base rates; power calculations revealed that 824 and 501 cases respectively would be required to detect a significant effect. Further, the SOS (Nicholls et al., 2007) lacks a strong evidence base, although it likely produced a more valid outcome measure than would have been achieved by reliance on whether electronic progress notes had been flagged by clinical staff as containing an adverse incident of each type.

3.5.1 Author's contribution

The candidate made a substantial contribution to designing the study and acquiring data and was responsible for data analysis. The candidate completed the first draft of the article and revised the manuscript in collaboration with co-authors. The candidate approved the final version for publication. This was deemed to represent 45% of total contributions.

3.6 Paper 6: Predictive validity of the START for unauthorised leave and substance abuse in a secure mental health setting (O'Shea & Dickens, 2015b)

As paper 5 was insufficiently powered to examine the ability of the START to inform risk assessment of unauthorised leave and substance abuse, this study extended the previous paper by investigating these outcomes in a sufficiently powered study (see 5.12 for full paper). This paper followed the same design as paper 5 and investigated the predictive ability of the START strength scores,

vulnerability scores, and relevant specific risk estimates, for the occurrence of unauthorised leave and substance abuse in the three months following START assessment (*N*=827). Neither strength nor vulnerability scores predicted the occurrence of substance abuse but the vulnerability score was a significant predictor of unauthorised leave. However, none of the individual items, in either their strength or vulnerability iterations, were significantly predictive of this outcome. In contrast, items 5 (self-care) and 8 (substance abuse) predicted the occurrence of substance abuse when scored both in terms of strengths and vulnerabilities. Both of the specific risk estimates predicted their corresponding outcomes, however the lack of relevant items in the START for these outcomes questions the contribution of these scores to the specific risk estimate formulation. PPVs based on the specific risk estimates were low (5.9%-8.1%), whilst the NPVs were very high (98.4%-99%). These results suggest that the START may be best able to identify those individuals at low risk of these outcomes and preclude the use of unnecessary restrictions. The low PPVs caution against intervening in those rated as elevated risk of these outcomes without further assessment.

3.6.1 Author's contribution

The candidate made a substantial contribution to designing the study and acquiring data. They also contributed to revising the first draft for intellectual contribution and approved the final version for publication. The candidate and her co-author agreed that they contributed equally to this paper.

3.7 Paper 7: Gender differences in the predictive efficacy of the START for aggression and self-harm (O'Shea & Dickens, 2015a)

This paper is the first stage of investigating the second half of research question 2 (i.e., whether the accuracy of the START is affected by sample characteristics and the implications of this for treatment planning; see Figure 1). The most robust finding from papers 1 and 3 is that the HCR-20 had superior performance among women compared with men. Therefore, this study aimed to establish if the START is similarly affected by gender using rocreg analysis (see 5.13 for full paper). This study followed the same design, and used data from the same sample of patients, as Paper 6. The predictive validity of the START

strength and vulnerability scores, and specific risk estimates for self-harm, suicide, and violence were examined for inpatient aggression and self-harm occurring in the three months following START assessment in a sample of 149 men and 51 women. Examination of the predictive ability of individual items was also examined to inform treatment targets.

Prediction of verbal aggression and self-harm/suicide by the vulnerability scale was significantly more accurate in women compared with men; whilst the remaining comparisons were not significantly different. AUC values were larger in women with the exception of physical aggression predicted by the strength scale. Importantly, based on AUC values, vulnerability scores and the self-harm specific risk estimate significantly predicted self-harm/suicide in women but none of the START components were significantly predictive of this outcome in men. However, being rated as moderate or high risk on the self-harm specific risk estimates significantly increased odds of men engaging in self harm by 7.10 and 13.01, respectively. The specific risk estimates showed greater improvements over scale scores in women relative to men, suggesting that clinicians are considering additional information not covered by the START when forming their risk estimates for this group. Examination of individual items revealed differences in predictive potency as a function of gender; impulse control was among the most predictive items for all outcomes in males, but none of the AUC values reached the threshold for a large effect size in this group. External triggers appeared more important in women and produced a large effect size for the prediction of verbal and any aggression.

A key limitation of this paper is that we were missing a large number of specific risk estimates for self-harm and suicide for men; this is possibly due to higher rates of self-harm in women (Nijman & Campo, 2002) leading clinicians to believe that it is unnecessary to routinely assess risk of this outcome among males. Further, the female sample was about a third of the size of the male sample due to fewer women in secure mental health care (Rutherford & Duggan, 2007).

3.7.1 Author's contribution

The candidate made a substantial contribution to designing the study and acquiring data and was responsible for data analysis. The first draft of the article was completed by the candidate who revised it for intellectual contribution in collaboration with her co-author. The final version was approved by the candidate. The candidate and her co-author made equal contributions to this paper.

3.8 Paper 8: Predictive validity of the HCR-20 for inpatient self-harm (O'Shea, Picchioni, Mason, Sugarman, & Dickens, 2014b)

This paper aimed to answer the pragmatic question of whether the HCR-20 predicts inpatient self-harm (see 5.14 for full paper). Answering this research question was not an explicit aim at the start of the research projects described in section 1.1 but developed in response to evidence from papers 2 and 5 that the START items do not appear to be of relevance to self-harm. Whilst, the HCR-20 is not intended to assess risk of self-harm, there is an overlap in the people who engage in self-harm and those who engage in outwardly directed aggressive behaviour (Nijman & Campo, 2002). Therefore, it may be reasonable to assume that at least some of the risk factors for self-harm and outwardly directed aggression are shared. Given that time and resources are limited in clinical practice it would be beneficial to determine if the HCR-20, which is the most commonly used risk assessment tool in medium secure units in England (Khiroya et al., 2009), can predict self-harm, and whether it performs equally in different clinical and demographic groups.

The HCR-20 total score, historical scale score, and risk-management scale score significantly predicted the occurrence of any self-harm, repeated self-harm, and severe self-harm (level 3 or above on the OAS; Yudofsky et al., 1986). Whilst rocreg analysis revealed only one significant difference in performance between men and women, a larger number of items were predictive of self-harm for the latter group. The HCR-20 showed poorest performance in those with organic diagnoses, consistent with the results of paper 2. There was no difference in the performance of the HCR-20 as a function of ethnicity, but there were a greater number of relevant items for the

Caucasian group. Similarly, a larger number of items were relevant for the younger, compared with older, group.

The findings of this study open up the possibility for clinicians to make a second summary judgment regarding the risk of self-harm occurring; supplementary self-harm specific risk factors, such as parental separation, depression, and self-derogation (Fliege et al., 2009; Klonsky & Muehlenkamp, 2007; Sourander et al., 2006), may improve the predictive efficacy of the HCR-20 for this outcome. Interventions to reduce self-harm in women and individuals with comorbid schizophrenia and personality disorder should aim to increase support when adjusting to new circumstances, whereas interventions in schizophrenia may be better targeted at reducing noncompliance.

3.8.1 Author's contribution

The candidate was responsible for acquiring and analysing data and completed the first draft of the paper. The candidate contributed to revising the manuscript in collaboration with co-authors, and approved the final version for publication. This was judged by the senior author on the paper to represent 40% of total contributions.

3.9 Item-outcome analysis

The examination of the predictive validity of individual HCR-20 and START items in papers 3, 4, 6, 7, and 8 indicated which factors are the most potent predictors of the various adverse outcomes across groups. Whilst some of the historical items on the HCR-20 may have high potency, they cannot easily be targeted by risk-management strategies. Therefore, the focus in the following section, which summarises the current evidence from the papers included in this thesis and presents the HCR-20 and START items that are relevant to the greatest number of individuals, is placed on the dynamic items. The purpose of this is to highlight factors that will likely contribute to a reduction in negative events if targeted by risk-management procedures in order to address the aim presented in section 1.4.

3.9.1 Aggression

C4 (impulsivity) was one of the strongest predictors of aggression among all groups studied, suggesting it should be one of the factors targeted for individuals deemed at risk of aggression that score highly in this domain. Similarly, S9 (impulse control in its strength iteration) was one of the most potent predictors of not engaging in aggressive behaviour for men. Both mental state in its strength iteration (S6) and major mental illness (C3) were among the most potent predictors for men and women, as were factors related to noncompliance/adherence (S14: medication adherence, S15: rule adherence, and R4: noncompliance with remediation attempts). R2 (exposure to destabilisers) was a strong predictor of aggression in individuals with co-morbid personality disorder and schizophrenia, and in those with organic disorders, developmental disorders, and intellectual disability. Similarly, in its strength iteration external triggers (S10) was predictive of non-engagement in aggression in women. Finally, lack of personal support (R3) appears to be an important risk factor for aggression in women and those aged under 40 years of age, whereas R5 (stress) was a particular potent predictor of aggression among those with intellectual disability.

3.9.2 Self-harm and suicide

V15 (rule adherence), V19 (coping), R2 (exposure to destabilisers), R1 (plans lack feasibility), and C4 (impulsivity) were the strongest predictors of self-harm in women, and for most diagnostic groups examined, suggesting that these items should be important treatment targets for individuals that score highly on them and are deemed at risk of self-harm. For those with schizophrenia, negative attitudes (C2) and non-compliance with remediation attempts (R4) appear to be of greater relevance. None of the HCR-20 or START vulnerability items were significant predictors of self-harm among men. This suggests that the most relevant risk factors for males fall outside of the HCR-20, and to some extent the START and that, therefore, these tools will not be useful in informing risk management strategies for self-harm among this population. S9 (impulse control) was the only strength item to predict non-engagement in this group. The lack of predictive ability for impulsivity when considered in terms of risks/vulnerabilities (i.e., C4 and V9) suggests that impulsivity itself is not a

potent risk factor for self-harm among males, but that the ability to control impulses may protect those with an elevated risk for self-harm against engaging in the behaviour.

3.9.3 Unauthorised leave

The START vulnerability score was found to significantly predict unauthorised leave in paper 7. However, none of the individual START items were predictive of this outcome. This therefore suggests that the most relevant risk factors for unauthorised leave are not contained within the START and this tool is unlikely to be of use in identifying treatment targets to reduce the occurrence of this outcome.

3.9.4 Substance abuse

Unsurprisingly, the START substance abuse item was predictive of this behaviour in both its strength and vulnerability iterations (S8 and V8); self-care (S5 and V5) was the only remaining item to predict substance abuse.

3.9.5 Risk-management strategies

There are a number of interventions that may be effective for targeting the above risk factors. One of the crucial tasks for mental health professionals will be deciding which interventions are likely to be beneficial in each individual case and to prioritise the areas that need targeting first. These may be those that are the highest risk (i.e., the factors identified above that are relevant to the individual in question on which they have high vulnerabilities or low strengths), causing greatest distress to the individual, or are a barrier to effective engagement and interventions. For example, when individuals have high vulnerabilities for non-compliance this may need targeting, for example by improving the therapeutic alliance and identifying barriers to treatment (Borum, Swartz, Swanson, & Wiseman, 2001), before other interventions may be effective.

The HCR-20 companion guide (Douglas et al., 2001) outlines a number of approaches that may be effective at targeting the risk factors contained within the tool. For example, dialectal behaviour therapy (DBT) strategies (behaviour analysis, distress tolerance, emotion regulation, relationships) are suggested as

useful treatment targets for impulsivity (McMain & Courbasson, 2001), whereas cognitive behavioural approaches and modelling are likely to be promising interventions for tackling negative attitudes (Muller-Isberner, 2001). Further, the companion guide suggests that when targeting plans that lack feasibility it may be fruitful to trial these ideas, as long as they do not put the patient or others at risk, and that the experience of trial and error may be a helpful learning experience for the individual (Belfrage & Fransson, 2001). Lack of personal support appears to be another important factor for aggression. Given the different forms that support can take (Hengeller, Schoenwald, Borduin, Rowland, & Cunningham, 1998), it is unlikely that any intervention would be universally beneficial and misapplications of support may be counterproductive. Augimeri (2001) suggests that it may be helpful to define support for the patient so that they can determine what assistance would be most beneficial and help them build these systems into their lives to develop long-term support systems.

There is a vast literature regarding the role of environmental factors and ward atmosphere on aggression (Gadon, Johnstone, & Cooke, 2006; Hallett, Huber, & Dickens, 2014) and the current evidence suggests that exposure to destabilisers may also be an important risk factor for self-harm in women. Staff skilled in de-escalation techniques and emotion regulation, separating or moving patients in the face of imminent risk, and using organised activities to divert attention may be fruitful management strategies for limiting the role of the ward environment (Bowers, 2014; Hallett et al., 2014; Webster, Eaves, & Halpin, 2001). Further, Stress Inoculation Training (Meichenbaum, 2007), which consists of three overlapping phases (conceptual education, skills acquisition and consolidation, and application), is perhaps one of the most well-known stress management interventions and may be beneficial in reducing the likelihood of adverse incidents occurring in response to exposure to destabilisers, as well reducing general levels of stress. Finally, motivational interviewing, which first identifies and develops discrepancies between actual and desired behaviours creating cognitive dissonance, and then increases self-efficacy in the possibility of change (Miller & Rollnick, 2002), has been shown to be effective in improving self-care and health outcomes in a number of domains, including substance abuse (Martins & McNeil, 2009), and may

therefore be a promising intervention for individuals with high vulnerabilities for both substance abuse and self-care more widely.

3.10 Overall critique common to studies

One limitation that is common across the primary research presented in this thesis is the reliance on routinely collected information. Whilst this approach has a number of advantages, as described in section 2.4, we were missing a large amount of information pertaining to ethnicity and the specific risk estimates. This resulted in rather crude classification of ethnicity into Caucasian and non-Caucasian for the purpose of analysis and may have obscured potential ethnic differences in performance. A second problem is that the individuals tasked with managing adverse incidents (i.e., the multidisciplinary clinical team) were the same individuals responsible for completing the risk assessment, or at least had knowledge of its outcome. ROC analysis cannot disentangle successful prevention from false prediction; hence, this likely contributed to the smaller AUC values obtained in these studies in comparison with studies where the risk assessments have been completed by the research team. However, the procedure used in these papers better represents what occurs in clinical practice, and therefore has greater ecological validity. Further, if such tools are reliable, which has repeatedly been claimed (e.g., Desmarais et al., 2012; Douglas et al., 2013; Wilson, Desmarais, Nicholls, & Brink, 2010; Wilson, Desmarais, Nicholls, Hart, & Brink, 2013), then the clinical teams should independently come to the same, or very similar, conclusions as the research team and implement risk-management strategies accordingly.

4   Summary and future directions

4.1   The use of the HCR-20 and the START in informing risk in mental health inpatients

Figure 1 outlines what was known about the efficacy of the HCR-20 and the START for informing assessment and management of risk among secure mental health inpatients prior to this thesis, the papers that comprise the thesis, and the important contribution this thesis has made to the literature. Prior to this thesis, evidence suggested that the HCR-20 could predict inpatient aggression, but this evidence was derived from mainly male populations; evidence regarding its performance among female samples and across different ethnic and diagnostic groups was limited and produced mixed results (see section 2.3.3.1). Evidence had emerged for the predictive ability of the START, but this was mainly limited to the prediction of aggression and self-harm and there were no existing systematic reviews of this literature (see section 2.3.3.2). The presented papers have expanded this knowledge, some using statistical methods that have not previously been used in this area, to clarify what populations, and for which outcomes, these tools have predictive efficacy. Conversely, they highlight a number of areas where these tools do not perform effectively and should, therefore, not be used to make important decisions regarding level of imposed restrictions. Further, the item-outcome analyses presented in section 3.9 provide practical advice for clinicians conducting risk assessment and outline management strategies that may be effective in reducing the occurrence of adverse events. The following sections elaborate on these findings.

4.1.1   The HCR-20

The primary research papers presented in this thesis (see section 3) suggests that the HCR-20 total can be considered a multivariate risk factor for inpatient aggression in the following groups, with some variation across groups in the potency of individual items (see section 3.9.1): 1) men and women, with superior efficacy among women, 2) individuals with and without intellectual disability, 3) individuals with schizophrenia and personality disorder, 4) Caucasians and non-Caucasians, and 5) older and younger individuals. For

individuals with developmental and organic disorders, the HCR-20 only appears to be a multivariate risk factor for verbal aggression, suggesting that it does not provide a useful assessment of risk for physical aggression in these populations. The total HCR-20 score appears to represent a multivariate risk factor for self-harm in women and individuals with schizophrenia and personality disorder (see section 3.8); however, it is typically not as potent a risk factor for self-harm as it is for aggression. The items within the tool do not appear to comprise a multivariate risk factor for self-harm in men, or for those with developmental and organic disorders. The above findings indicate that it would be inappropriate to use the HCR-20 to inform decision making and management of aggression and self-harm for those with developmental and organic disorders, or to implement risk management strategies targeting self-harm among men. Doing so may lead to consequences as described in section 1.3 that can be costly for the patient themselves, staff, family members, organisations, and the general public.

4.1.2 The START

Paper 5 suggested that the START strength and vulnerability scales are multivariate risk factors for aggression; however, more fine-grained analysis of gender differences in paper 7 indicated that the vulnerability scale may only constitute a risk factor for this outcome in women. The specific risk estimate for violence is predictive of aggression in both groups and adds incrementally to prediction achieved by scores alone. Initial examination of the START's predictive efficacy in paper 5 suggested that the scores were not informative regarding risk of self-harm; however, paper 7 revealed that the vulnerability scale is a multivariate risk factor for this outcome in women only, whereas there is some evidence that the specific risk estimate has relevance for both groups. Despite this, the HCR-20 items, which are not intended to predict the occurrence of self-harm, achieve similar levels of predictive efficacy for this outcome than that achieved by the START (see section 3.8).

There is no evidence that the START strength and vulnerability total scores have predictive ability for either victimisation or substance abuse, but some of the individual START items are of relevance to the latter outcome (see Section

3.9.4) and both outcomes were predicted by their corresponding risk estimates. The START strength scale, but not the vulnerability scale, appears to be a multivariate risk factor for engaging in self-neglect (see section 3.5); however, the corresponding specific risk estimate does not predict this outcome. In contrast, the START vulnerability score, but not the strength score, predicted the occurrence of unauthorised leave, suggesting that it may constitute a multivariate risk factor for this outcome; however, none of the individual items predicted its occurrence. The specific risk estimate for unauthorised leave was predictive of its corresponding outcome.

The above findings suggest that the START does not contain risk factors that are relevant to all of the outcomes it intends to predict and that the relevance of the items is largely limited to aggressive outcomes. This casts considerable doubt on its ability to inform risk management strategies for the remaining outcomes and suggests that it would be inappropriate to base such decisions on the START. The superior predictive efficacy of the specific risk estimates, with the exception of that for self-neglect, indicates that clinicians are able to form reliable risk estimates for the majority of outcomes; however, it is unclear to what extent these are actually based on consideration of the START items. There is some evidence from a recent study that clinicians appear to give most weight to recent adverse behaviours when forming specific risk estimates, rather than scores on the START items (O'Shea & Dickens, 2016).

4.2 Directions for future research

4.2.1 Further investigation of predictive validity of the HCR-20 and the START

Despite the contributions of this thesis outlined in the previous sections, research on risk assessment and management of secure mental health inpatients is, to some extent, still in its infancy and further research is needed. There are a number of outstanding research questions in relation to the performance of the HCR-20 and the START; more complex study designs that allow for the inclusion of risk-management strategies as moderators of the relationship between risk assessment scores and adverse outcomes may help

disentangle whether observed poor performance is due to false prediction or successful prevention.

Section 3.9 outlined the most-potent predictors of the various outcomes included in this thesis, with the exception of victimisation and self-neglect. Given that neither the strength nor vulnerability scale significantly predicted victimisation, it is unlikely that the START contains particularly potent items for victimisation and future research efforts regarding this outcome are perhaps best directed towards the identification of additional factors. However, as the START strength scale was a significant predictor of self-neglect, future research should investigate the potency of the individual items for this outcome. Further investigations of the differential predictive validity of the START is also warranted, particularly across diagnostic groups due to the lack of efficacy of the HCR-20 among individuals with developmental and organic disorders.

Paper 8 indicated that the HCR-20 scores demonstrated similar, if not superior, predictive efficacy for self-harm than that achieved by START scores. Further, it is more established than the START both in terms of its evidence base and its incorporation into clinical practice. Therefore, it would be pragmatic to consider whether clinicians can form accurate summary judgments regarding the risk of this outcome based on the consideration of the HCR-20 items; the provision of additional information concerning empirical links between the factors contained in the HCR-20 and self-harming behaviours may facilitate these judgements. Given that this thesis has indicated extremely limited predictive ability of the START for outcomes beyond aggression and self-harm, if these judgments prove to be effective it may be more practical for clinicians to invest their limited time and resources in risk assessment and management of both aggression and self-harm based on the HCR-20, until there has been some refinement of the START items.

4.2.2 Determining the effectiveness of the HCR-20 and the START for informing active risk-management

The ultimate aim of risk assessment is to manage identified risks; hence, from a prevention-based risk management perspective, risk assessment can be considered successful when there is a reduction in rates of adverse events

associated with assessment procedures (Douglas & Kropp, 2002). This does not negate the need to examine the predictive efficacy of risk assessment tools, which has been the focus of the papers presented in this thesis; it is important to determine that they contain causal risk factors (i.e., factors that change the likelihood of the outcome occurring when manipulated; Kraemer et al., 1997) if they are to assist in active management of risk. However, such research should exist to further preventative strategies (Douglas & Kropp, 2002). Therefore, it is important to establish 1) whether HCR-20 and START scores and risk judgments can change over time, and 2) whether associated changes are linked to a reduction in risk behaviour. However, to date, the limited research in this area (Douglas & Kropp, 2002) has instead focused on the more general question of whether risk assessment with these tools leads to a reduction in adverse events, with mixed findings, or has examined changes in scores without investigating associated changes in risk behaviour.

Longer treatment times have been found to be associated with reduced scores on the clinical and risk-management scales of the HCR-20 (Belfrage & Douglas, 2002; Sheldon & Gallagher, 2010) and direct examination of this has shown that mean scores reduce across repeated HCR-20 assessments (Morrissey et al., 2007; O'Shea & Dickens, 2015c; Olsson, Strand, Kristiansen, Sjöling, & Asplund, 2013). This suggests that at least some of the HCR-20 risk factors are variable (i.e., can change). However, mean changes in scores are very small (less than 0.5 points per iteration) and O'Shea and Dickens (2015c) found that only 3% of patients demonstrated rates of change that would be considered both reliable and clinically significant.

With regard to variation in incidents, Crocker et al. (2008) found a decrease in the number of incident reports during the 15 months following the implementation of the START at a risk-management and rehabilitation unit. However, they did not examine changes in incidents pre- and post-implementation and the implementation coincided with significant changes to admission criteria and the care program that may have contributed to a decrease in incidents. In contrast, Troquete et al. (2013) found that intervention following assessment with the Dutch version of the START ('t Lam, Lancel, & Hildebrand, 2009) did not reduce violent and criminal behaviour in the

community, relative to a treatment as usual group that received no structured risk assessment. However, the number of incidents reduced between baseline and follow-up for both groups and there was poor fidelity to study protocol in the intervention group.

Very few researchers have concurrently examined changes in risk factors and adverse incidents which, as previously stated, is required to determine whether risk assessment tools can be classified as containing causal risk factors. Incidents of aggression in a maximum security prison decreased following the introduction of the HCR-20 but there was no significant change in risk factors as measured by the HCR-20 (Belfrage, Fransson, & Strand, 2004). However, change in incidents and scores were only examined at a group level, which may have obscured changes in individuals' risk. Whittington et al. (2014) found a 10% reduction in vulnerabilities as measured by repeated START assessments over a period of 27 months; however, there was also a 15% reduction in strengths. A ten-point increase in vulnerability scores was associated with increased risk of engaging in subsequent violence (OR=3.10); however, this represents a dramatic change in scores which may be unlikely to commonly occur in practice.

The weight of the above evidence may suggest that risk factors contained in the HCR-20 are relatively fixed and/or unrelated to aggression, or more conservatively are not the most potent predictors of this outcome. Alternatively, study periods may have been of insufficient length to allow for meaningful change, selected interventions may not have been effective at reducing risk, or teams rating assessments may be insensitive to change. However, the evidence was mixed, difficult to interpret, contained confounding factors (e.g., poor fidelity to study design and concurrent change to care), and some were conducted outside of the secure mental health setting. Further, there is a dearth of literature regarding changes in rates of adverse outcomes beyond aggression and criminal behaviour (e.g., self-harm and suicide). Therefore, further research efforts are needed in this area to provide clearer evidence as to whether the HCR-20 and the START contain causal risk factors for those outcomes and patient groups where they have predictive efficacy, as demonstrated by this

thesis, and ultimately whether these tools can move beyond passive prediction and assist in the implementation of effective management strategies.

It may therefore be fruitful to investigate if strategies targeting the most potent predictors of adverse outcomes for a given population, such as those identified in section 3.9, are effective in reducing the occurrence of adverse events. The optimal design for this may involve the following steps: 1) measuring the occurrence of adverse incidents, using standardised methods, prior to any risk assessment taking place, 2) risk assessment conducted by independent members of clinical team with established inter-rater reliability, 3) implementation of risk management strategies, and 4) subsequent risk assessment and collection of adverse outcome data following a sufficient time period for interventions to have been effective. This would allow researchers to determine if: 1) the HCR-20 and the START scores reduce following interventions and therefore contain variable risk factors, and 2) whether reductions in these factors are linked to reduction in adverse outcomes, and hence whether such factors can be considered causal. This process could then be iterated to examine changes in levels of risk over the course of inpatient treatment.

4.2.3 Identification of additional risk and protective factors

This thesis highlighted a number of shortcomings of current risk assessment methods. Future research should aim to identify relevant risk and protective factors for: 1) self-harm and suicidal behaviours among male secure mental health inpatients, 2) aggression and self-harm/suicide for individuals with organic and developmental disorders, and 3) unauthorised leave, substance abuse, and victimisation for the whole range of secure mental health inpatients. Despite the lack of strong evidence for the HCR-20 and START as multivariate risk factors for the above outcomes and populations, there are some individual items (as discussed in section 3.9) that have demonstrated predictive efficacy that should be fruitful treatment targets while additional relevant factors are identified (e.g., exposure to destabilisers as a predictor of aggression for those with organic and developmental disorders). Once identified, these factors

should be used to develop new or existing risk assessment schemes and their efficacy should be tested.

4.2.4 Investigations of recent developments in risk assessment

There have been a number of further developments to risk assessment tools in recent years that should be investigated by future research. As stated in the introduction, since data collection for the HCR-20 project finished, version 3 of the HCR-20 has been developed (Douglas et al., 2013). Early evidence suggests that performance of the two tools is very similar (e.g., Strub et al., 2014) and there is some evidence that version 3 does not have improved accuracy relative to version 2 (Coid, Kallis, Doyle, Shaw, & Ullrich, 2015). Further research is needed to clarify whether version 3 improves accuracy, particularly for those groups where version 2 did not have predictive efficacy. Further, de Vogel et al. (2012) developed the Female Additional Manual (FAM) which contains additional female-specific risk factors and coding instructions for use with version 2 of the HCR-20 in female populations; it has since been modified for use with version 3 (de Vogel, de Vries Robbe, Van Kalmthout, & Place, 2014). This development seems to have been largely driven by theoretical assumptions that aggressive and violent behaviour by women is associated with their unique experiences as women (as discussed in section 1.3.1) and that adequate risk assessment requires an appreciation of such issues (i.e., the 'gendered persepctive'; see Nicholls et al., 2004), rather than by empirical evidence that current risk assessment methods are inadequate for women which appears not to be the case in light of the presented papers. However, that does not preclude the possibility that the FAM contains more relevant and potent risk factors than the standard HCR-20 for aggression and violence among women and is therefore better able to assist in the risk assessment and management of this population; this should be examined in future research. Similarly, additional guidelines are available for scoring of items on version 2 of the HCR-20 for individuals with intellectual disability (Boer, Frize, Pappas, Morrissey, & Lindsay, 2010); there is some evidence that their use results in slightly larger AUC values than those obtained from the HCR-20 under standard scoring instructions (Verbrugge, Goodman-Delahunty, & Frize, 2011) but they have not been extensively researched.

67

4.3 Conclusion

This thesis is the result of two related projects regarding the use of the HCR-20 and the START to inform risk assessment and management of secure mental health inpatients. The presented papers have used considerably novel methods, high quality, routinely collected data, and larger sample sizes than have previously been reported in this area. They also represent one of the largest bodies of research in this area (in terms of number of papers and sample sizes) that has been conducted by independent researchers; the majority of the available literature has been conducted by the tools' authors, or individuals affiliated with them, which has been shown to produce larger effect sizes (Singh, Grann, & Fazel, 2013). As demonstrated in Figure 1, the papers that comprise this thesis represent significant advancements in our knowledge of the performance of structured professional judgement tools in the risk assessment of secure mental health patients. However, section 4.2 highlights the necessity for future research in this area. It is recommended that the methods discussed and utilised in this thesis should be adopted into routine clinical practice to ensure the availability of high quality, reliable data that can be used both to feedback into practice and to provide a resource for future research. Further, the possible directions for future research should shift the focus towards thoroughly investigating the impact of existing risk assessment tools and whether they can reduce risk and subsequent adverse incidents rather than developing additional tools that may not be required, or represent an improvement over existing methods.

5   Appendices

5.1   Appendix A: Related publications, citations and other dissemination

Citations

Paper 1: Moderators of the predictive efficacy of the HCR-20: Systematic review and meta-analysis (O'Shea et al., 2013)

Arai, K., Takano, A., Nagata, T., & Hirabayashi, N. (2016). Predictive accuracy of the Historical-Clinical-Risk Management-20 for violence in forensic psychiatric wards in Japan. *Criminal behaviour and mental health*.

Čėsnienė I., Klimukienė V. (2014). Smurtinio elgesio rizikos vertinimas: HCR-20 prognostinis validumas už smurtinius nusikaltimus teistų asmenų imtyje. *Kriminologijos studijos, 2*, 138–154.

de Vogel, V., de Vries Robbé, M., van Kalmthout, W., & Place, C. (2014). Female Additional Manual: Additional guidelines to the HCR-20v 3for assessing risk of violence in women (English version). Van der Hoeven Kliniek.

de Vogel, V. & Nicholls, T. L. (2016). Gender Matters: An Introduction to the Special Issues on Women and Girls. *International Journal of Forensic Mental Health, 15*(1), 1-25.

de Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. (2014). Changes in Dynamic Risk and Protective Factors for Violence During Inpatient Forensic Psychiatric Treatment: Predicting Reductions in Postdischarge Community Recidivism. *Law and Human Behavior*

de Vries Robbé, M., de Vogel, V., Wever, E. C., Douglas, K. S., & Nijman, H. L. (2016). Risk and protective factors for inpatient aggression. *Criminal Justice and Behavior*

Douglas, K. S. (2014). Version 3 of the Historical-Clinical-Risk Management-20 (HCR-20v3): Relevance to violence risk assessment and management in forensic conditional release contexts. *Behavioural Sciences and the Law, 32*, 557-576.

Douglas, K., Hart, S., Webster, C., & Belfrage, H. (2013). HCR-20: Assessing risk for violence (Version 3). Burnaby, BC: Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K. S., Shaffer, C., Blanchard, A. J. E., Guy, L. S., Reeves, K. A., & Weir, J. (2014). HCR-20 violence risk assessment scheme: Overview and annotated bibliography (current up to January 1, 2014).

Green, D., Belfi, B., Griswold, H., Schreiber, J. M., Prentky, R., & Kunz, M. (2014). Factors associated with recommitment of NGRI acquittees to forensic hospital. *Behavioural Sciences and the Law, 32*, 608-626.

Geraghty, K. A., & Woodhams, J. (2015). The predictive validity of risk assessment tools for female offenders: a systematic review. *Aggression and Violent Behavior, 21*, 25-38.

Grimbos, T., Penney, S. R., Fernane, S., Prosser, A., Ray, I., & Simpson, A. I. (2016). Gender Comparisons in a Forensic Sample: Patient Profiles and HCR-20: V2 Reliability and Item Utility. *International Journal of Forensic Mental Health,* 1-13.

Harris, G. T., & Rice, M. E. (2015). Progress in violence risk assessment and communication: Hypothesis versus evidence. *Behavioural Sciences and the Law, 33*, 128-145.

Morrissey, C., Beeley, C., & Milton, J. (2013). Longitudinal HCR-20 scores in a high-secure psychiatric hospital. Criminal Behaviour and Mental Health

Murphy, D. (2013). Risk assessment of offenders with an autism spectrum disorder. *Journal of Intellectual Disabilities and Offending Behaviour, 4*, 33-41.

Penney, S. R., McMaster, R., & Wilkie, T. (2014) Multirater reliability of the Historical, Clinical, and Risk Management-20. *Assessment, 21*, 15-27.

Sedgwick, O., Young, S., Das, M., & Kumari, V. (2016). Objective predictors of outcome in forensic mental health services—a systematic review. *CNS spectrums*, 1-15.

Sen, P., Lindsey, S., Chatterjee, N., Rama-Iyer, R., & Picchioni, M. (2015). An audit of the quality of HCR-20 violence risk assessments in a low secure service. *Journal of Psychiatric Intensive Care*, 1-9.

Simpson, A. I., Penney, S. R., Fernane, S., & Wilkie, T. (2015). The impact of structured decision making on absconding by forensic psychiatric patients: results from an AB design study. *BMC psychiatry, 15*(1), 103.

Troquete, N. A. C., van den Brink, R. H. S., Beintema, H., Mulder, T., van Os, T. W. D. P., Schoevers, R. A., & Wiersma, D. (2015). Predictive Validity of the Short-Term Assessment of Risk and Treatability for Violent Behavior in Outpatient Forensic Psychiatric Patients. *Psychological Assessment, 27*(2), 377-391.

Viljoen, S., Nicholls, T. L., Roesch, R., Gagnon, N., Douglas, K., & Brink, J. (2016). Exploring Gender Differences in the Utility of Strength-Based Risk Assessment Measures. *International Journal of Forensic Mental Health*, 1-15.

Vojt, G. (2014). *Implementation of violence risk assessments into forensic psychiatric care in Scotland.* (Unpublished doctoral thesis). University of Edinburgh, Scotland.

Paper 2: Short-Term Assessment of Risk and Treatability (START): Systematic review and meta-analysis (O'Shea & Dickens, 2014)

Alderman, N., Major, G., & Brooks, J. (2016). What can structured professional judgement tools contribute to management of neurobehavioural disability? Predictive validity of the Short-Term Assessment of Risk and Treatability (START) in acquired brain injury. *Neuropsychological rehabilitation*, 1-18

Berman, N.C., Sullivan, A., Wilhelm, S., & Cohen, I. G. (2015). Effect of a Legal Prime on Clinician's Assessment of Suicide Risk. *Death studies*

Blanchard, A. J., Shaffer, C. S., & Douglas, K. S. (2016). 18 Decision Support Tools in the Evaluation of Risk for Violence. In. P. M. Kleespies. *The Oxford Handbook of Behavioral Emergencies and Crises* (pp. 262-288). Oxford, UK: Oxford University Press.

Brown, J., & Singh, J. P. (2014). Forensic risk assessment: a beginner's guide. Archives of Forensic Psychology, 1, 49-59.

Cook, A. N., Moulden, H. M., Mamak, M., Lalani, S., Messina, K., & Chaimowitz, G. (2016). Validating the Hamilton Anatomy of Risk Management–Forensic Version and the Aggressive Incidents Scale. *Assessment*

Dickens, G. L. (2015). Re-focusing risk assessment in forensic mental health nursing. *Journal of psychiatric and mental health nursing*, *22*(7), 461-462.

Dickens, G. L., & Doyle, M. (2016). Mentally disordered firesetters in secure mental health care. In. R. M. Doley, G. L. Dickens, & T. A. Gannon. *The psychology of arson: A practical guide to understanding and manging deliberate firesetters* (pp. 260-275). Abingdon, UK: Routledge.

Gildberg, F. A., Fristed, P., Makransky, G., Moeller, E. H., Nielsen, L. D., & Bradley, S. K. (2015). As Time Goes by: Reasons and characteristics of prolonged episodes of mechanical restraint in forensic psychiatry. Journal of forensic nursing, 11(1), 41-50.

Kivisto, A. J. (2015). Violence risk assessment and management in outpatient clinical practice. *Journal of clinical psychology, 72*(4), 329-349.

Lowder, E. M. (2015). Models of protection against Recidivism in Adults with Mental Illnesses. (Masters thesis). North Carolina State University.

Petersen, K. L. (2015). The transition of British Columbia's tertiary psychiatric services from hospital to community: consideration of clinical, psychosocial, and transinstitutionalization outcomes. (Doctoral thesis). University of British Columbia.

Singh, J. P. (2016). International Perspectives on Forensic Risk Assessment: Measuring Use, Perceived Utility, and Research Quality (Doctoral thesis). Universität Konstanz.

Viljoen, S. (2014). *Examining the utility of strengths and protective factors in violence risk assessment measures in a tertiary civil psychiatric population.* (Doctoral dissertation). Simon Fraser University.

Viljoen, J. L., Gray, A. L., & Barone, C. (2015). Assessing Risk for Violence and Offending in Adolescents. In R. Jackson, & R. Roesch. *Learning Forensic Assessment: Research and Practice*, (pp. 357-388). New York; Routledge.

Viljoen, S., Nicholls, T. L., Roesch, R., Gagnon, N., Douglas, K., & Brink, J. (2016). Exploring Gender Differences in the Utility of Strength-Based

Risk Assessment Measures. *International Journal of Forensic Mental Health*, 1-15.

Paper 3: Differential predictive validity of the HCR-20 for inpatient aggression (O'Shea et al., 2014a)

Alderman, N., Major, G., & Brooks, J. (2016). What can structured professional judgement tools contribute to management of neurobehavioural disability? Predictive validity of the Short-Term Assessment of Risk and Treatability (START) in acquired brain injury. *Neuropsychological rehabilitation*, 1-18.

Arai, K., Takano, A., Nagata, T., & Hirabayashi, N. (2016). Predictive accuracy of the Historical-Clinical-Risk Management-20 for violence in forensic psychiatric wards in Japan. *Criminal behaviour and mental health*.

Green, D., Schneider, M., Griswold, H., Belfi, B., Herrera, M., & DeBlasi, A. (2016). A Comparison of the HCR-20V3 Among Male and Female Insanity Acquittees: A Retrospective File Study. *International Journal of Forensic Mental Health*, *15*(1), 48-64.

Ireland, J. L., Priday, L. J., Ireland, C. A., Chu, S., Kilcoyne, J., & Mulligan, C. (2016). Predicting hospital aggression in secure psychiatric care. *British Journal of Psychiatry Open*, *2*(1), 96-100.

Knezevic, V., Mitrovic, D., Drezgic-Vukic, S., Knezevic, J., Ivezic, A., Siladji-Mladenovic, D., & Golubovic, B. (2015). Prevalence and correlates of aggression and hostility in hospitalized schizophrenic patients. *Journal of interpersonal violence*

Rampersad, K., Berding, R., Esq, M. S., & Swift, R. (2016). Overview of Psychiatric Patient Safety Issues. *URBAN MEDICINE*, 65.

Strub, D. S., Douglas, K. S., & Nicholls, T. L. (2016). Violence Risk Assessment of Civil Psychiatric Patients with the HCR-20: Does Gender Matter? *International Journal of Forensic Mental Health*, *15*(1), 81-96.

Viljoen, S., Nicholls, T. L., Roesch, R., Gagnon, N., Douglas, K., & Brink, J. (2016). Exploring Gender Differences in the Utility of Strength-Based Risk Assessment Measures. *International Journal of Forensic Mental Health*, 1-15.

Paper 4: The effect of intellectual disability on accuracy of the HCR-20 (O'Shea et al., 2015)

Pouls, C., & Jeandarme, I. (2015). Risk Assessment and Risk Management in Offenders with Intellectual Disabilities: Are We There Yet?. *Journal of Mental Health Research in Intellectual Disabilities*, *8*(3-4), 213-236.

Paper 5: The predictive validity of the Short-Term Assessment of Risk and Treatability (START) for multiple adverse outcomes in a secure psychiatric inpatient setting (O'Shea et al., 2016)

Dickens, G. L. (2015). Re-focusing risk assessment in forensic mental health nursing. *Journal of psychiatric and mental health nursing*, *22*(7), 461-462.

Font Pujol, J. (2016). Violència i psiquiatria: validació de l'Escala de Valoració de Risc d'Agressivitat (EVRA). (Doctoral thesis). Universitat de Girona.

Petersen, K. L. (2015). The transition of British Columbia's tertiary psychiatric services from hospital to community: consideration of clinical, psychosocial, and transinstitutionalization outcomes. (Doctoral thesis). University of British Columbia.

Singh, J. P. (2016). International Perspectives on Forensic Risk Assessment: Measuring Use, Perceived Utility, and Research Quality (Doctoral thesis). Universität Konstanz.

Viljoen, S., Nicholls, T. L., Roesch, R., Gagnon, N., Douglas, K., & Brink, J. (2016). Exploring Gender Differences in the Utility of Strength-Based Risk Assessment Measures. *International Journal of Forensic Mental Health*, 1-15.

Paper 6: Predictive validity of the START for unauthorized leave and substance abuse in a secure mental health setting (O'Shea & Dickens, 2015b)

Dickens, G. L. (2015). Re-focusing risk assessment in forensic mental health nursing. *Journal of psychiatric and mental health nursing, 22*(7), 461-462.

Paper 7: Gender differences in the predictive efficacy of the START for aggression and self-harm (O'Shea & Dickens, 2015a)

de Vogel, V. & Nicholls, T. L. (2016). Gender Matters: An Introduction to the Special Issues on Women and Girls. *International Journal of Forensic Mental Health, 15*(1), 1-25.

Viljoen, S., Nicholls, T. L., Roesch, R., Gagnon, N., Douglas, K., & Brink, J. (2016). Exploring Gender Differences in the Utility of Strength-Based Risk Assessment Measures. *International Journal of Forensic Mental Health*, 1-15.

Paper 8: Predictive validity of the HCR-20 for inpatient self-harm (O'Shea et al., 2014b)

Rampersad, K., Berding, R., Esq, M. S., & Swift, R. (2016). Overview of Psychiatric Patient Safety Issues. *URBAN MEDICINE*, 65.

Selenius, H., Leppänen Östman, S., & Strand, S. (2016). Self-harm as a risk factor for inpatient aggression among women admitted to forensic psychiatric care. *Nordic journal of psychiatry*, 1-7.

Strub, D. S., Douglas, K. S., & Nicholls, T. L. (2016). Violence Risk Assessment of Civil Psychiatric Patients with the HCR-20: Does Gender Matter? *International Journal of Forensic Mental Health*, *15*(1), 81-96.

Related conference presentations

Dickens, G. and O'Shea, L. (2012) Predictive validity of HCR-20 among heterogeneous groups of secure psychiatric inpatients: results of an audit. Paper presented to: Third International Conference on Violence in the Health Sector: Linking Local Initiatives with Global Learning, Vancouver, Canada, 24-26 October 2012.

O'Shea, L. E., Mitchell, A. E., Picchioni, M. M. & Dickens, G. L. (2013, June). Moderators of the predictive efficacy of the Historical, Clinical and Risk Management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis. In G. L. Dickens (Chair), Risk Prediction, Planning and Beyond: Structured Professional Judgment Approaches to Risk in

Forensic Settings. Symposium conducted at the BPS Division of Forensic Psychology Annual Conference, Belfast

Dickens, G. L., O'Shea. L.E., Picchioni, M.M., Mason, F.L., & Sugarman, P.A. (2013, June). All things being equal? Predictive validity of the HCR-20 among heterogeneous groups of secure psychiatric inpatients. In G. L. Dickens (Chair), Risk Prediction, Planning and Beyond: Structured Professional Judgment Approaches to Risk in Forensic Settings. Symposium conducted at the BPS Division of Forensic Psychology Annual Conference, Belfast

O'Shea, L. E., Mitchell, A. E., Picchioni, M. M. & Dickens, G. L. (2013, August). Moderators of the predictive efficacy of the Historical, Clinical and Risk Management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis. Paper presented at CPD event at St. Andrew's, Northampton.

Dickens, G. L., O'Shea. L.E., Picchioni, M.M., Mason, F.L., & Sugarman, P.A. (2013, August). Equality for all? Differential predictive validity of HCR-20 for aggression in psychiatric care. Paper presented at CPD event at St. Andrew's, Northampton.

Dickens, G. L., O'Shea. L.E., Mason, F.L., & Picchioni, M.M. (2013, October). Predictive validity of the HCR-20 among heterogeneous groups of secure psychiatric inpatients. Paper presented at the 8th European Congress on Violence in Clinical Psychiatry, Ghent.

O'Shea, L. E., Picchioni, M.M., Mason, F. L., Sugarman, P.A., & Dickens, G. L. (2014, June). Differential predictive validity of the HCR-20 for inpatient aggression. Paper presented at the Broadmoor Quarterly Seminar, Berkshire.

O'Shea, L.E., & Dickens, G. L. (2014, June). Short-Term Assessment of Risk and Treatability (START): Systematic Review and Meta-Analysis. Paper presented at CPD event at St. Andrew's, Northampton.

Dickens, G.L., O'Shea, L.E., & Picchioni, M.M. (2014, June). The predictive validity of the Short-Term Assessment of Risk and Treatability (START). Paper presented at CPD event at St. Andrew's, Northampton.

O'Shea, L. E., Picchioni, M.M., Mason, F. L., Sugarman, P.A., & Dickens, G. L. (2014, June). All things being equal? Predictive validity of the HCR-20

among heterogeneous groups of secure psychiatric inpatients. Paper presented at the 14th Annual Meeting of the International Association of Forensic Mental Health Services, Toronto.

O'Shea, L.E., & Dickens, G. L. (2014, June). Short-Term Assessment of Risk and Treatability (START): Systematic Review and Meta-Analysis. In G. L. Dickens (Chair), START to make a difference: Risk prediction, assessment and translation in the UK. Symposium conducted at the 14th Annual Meeting of the International Association of Forensic Mental Health Services, Toronto.

Dickens, G.L., O'Shea, L.E., & Picchioni, M.M. (2014, June). The predictive validity of the Short-Term Assessment of Risk and Treatability (START). In G. L. Dickens (Chair), START to make a difference: Risk prediction, assessment and translation in the UK. Symposium conducted at the 14th Annual Meeting of the International Association of Forensic Mental Health Services, Toronto.

Related poster presentations

O'Shea, L. E., Mason, F. L., Picchioni, M.M., & Dickens, G. L. (2013, July). All things being equal? Predictive validity of the HCR-20 among heterogeneous groups of secure psychiatric inpatients. Poster presented at the Royal College of Psychiatrists' International Congress, Edinburgh.

O'Shea, L. E., & Dickens, G. L. (2013, October). Short-Term Assessment of Risk and Treatability (START): Systematic Review and Meta-analysis. Poster presented at the Faculty of General Adult Psychiatry Annual Conference, Manchester.

O'Shea, L.E., & Dickens, G. L. (2014, March). Short-Term Assessment of Risk and Treatability (START): Systematic Review and Meta-Analysis. Poster presented at the Faculty of Forensic Psychiatry Annual Conference, Belfast.

O'Shea, L. E., Picchioni, M.M., Mason, F. L., & Dickens, G. L. (2014, March). All things being equal? Predictive validity of the HCR-20 among heterogeneous groups of secure psychiatric inpatients. Poster presented at the Faculty of Forensic Psychiatry Annual Conference, Belfast.

O'Shea, L. E., Picchioni, M.M., Mason, F. L., & Dickens, G. L. (2014, March). All
things being equal? Predictive validity of the HCR-20 among
heterogeneous groups of secure psychiatric inpatients. Poster presented
at the 22nd European Congress of Psychiatry, Munich

O'Shea, L.E., & Dickens, G.L. (2014, April). The predictive validity of the Short-
Term Assessment of Risk and Treatability (START). Poster presented at
the 11th National Conference of Research in Forensic Mental Health
Services, London.

O'Shea, L.E., Mason, F.L., Picchioni, M.M., & Dickens, G.L. (2014, June).
Predictive validity of the HCR-20 for inpatient aggression: The effect of
intellectual disability on accuracy. Poster presented at the 14th Annual
Meeting of the International Association of Forensic Mental Health
Services, Toronto.

Gunenc, C., O'Shea, L.E., & Dickens, G.L. (2014, December). Prevalence and
predictors of verbal aggression in a secure mental health service: use of
the HCR-20. Poster presented at the BPS Division of Clinical Psychology
Annual Conference, Glasgow

O'Shea, L.E., Picchioni, M.M., & Dickens, G.L. (2014, December). The
predictive validity of the Short-Term Assessment of Risk and Treatability
(START) for multiple adverse outcomes. Poster presented at the BPS
Division of Clinical Psychology Annual Conference, Glasgow

Related publications

Dickens, G. & O'Shea, L. (2012). Predictive validity of the HCR-20 among
heterogeneous groups of secure psychiatric inpatients: Results of an
audit. In. I. Needham, K. McKenna, M. Kingma, & N. Oud (Eds.).
Violence in the health sector: Linking local initiatives with global learning
(pp.402-403). Amsterdam: Kavanah

Dickens, G. L., & O'Shea, L. E. (2015). How short should short-term risk
assessment be? Determining the optimum interval for START
reassessment in a secure mental health service. *Journal of psychiatric
and mental health nursing, 22*(6), 397-406

Gunenc, C., O'Shea, L. E., & Dickens, G. L. (2015). Prevalence and predictors of verbal aggression in a secure mental health service: Use of the HCR-20. *International journal of mental health nursing*, *24*(4), 314-323.

O'Shea, L., & Dickens, G. (2015). Contribution of Protective Factors Assessment to Risk Prediction: Systematic Review and Meta-analysis.*European Psychiatry*, *30*, 211.

O'Shea, L. E., & Dickens, G. L. (2015). The HCR-20 as a measure of reliable and clinically significant change in violence risk among secure psychiatric inpatients. *Comprehensive psychiatry*, *62*, 132-140.

O'Shea, L. E., & Dickens, G. L. (2016). Role of assessment components and recent adverse outcomes in risk estimation and prediction: Use of the Short Term Assessment of Risk and Treatability (START) in an adult secure inpatient mental health service. *Psychiatry research*, *240*, 398-405.

O'Shea, L., Mason, F., Picchioni, M., & Dickens, G. (2014). EPA-0442 – All things being equal? Predictive validity of the HCR-20 among heterogeneous groups of secure psychiatric inpatients. European Psychiatry, 29, Supplement 1

O'Shea, L. E., Thaker, D. K., Picchioni, M. M., Mason, F. L., Knight, C., & Dickens, G. L. (2015). Predictive validity of the HCR-20 for violent and non-violent sexual behaviour in a secure mental health service. *Criminal Behaviour and Mental Health*.

Roychowdhury, A., Natarajan, M., O'Shea, L., & Dickens, G. (2015) Clinical risk assessment in secure care. In P. Sugarman, M. Picchioni, & G. Dickens. *Handbook of Secure Care.* London, UK: RCPSych Publications

Awards and acknowledgements

- Third International Conference on Violence in the Health Sector (2012): Best paper (Paper 3)
- Faculty of General Adult Psychiatry Annual Conference, Manchester (2013): Bonus poster prize (Paper 2)
- 22nd European Congress of Psychiatry, Munich (2014): Best poster prize (Paper 3)

- University of York Centre for reviews and dissemination (CRD) has determined that the HCR-20 and START meta-analyses (Papers 1 and 2) meet the DARE scientific quality criteria for a systematic review

5.2 Appendix B: OAS/MOAS scoring criteria (Kay et al., 1988; Yudofsky et al., 1986) – Copyright holder Stuart C. Yudofsky

This has been removed by the author of this thesis for copyright reasons

5.3 Appendix C: SOS scoring criteria (Nicholls et al., 2007) - Copyright holders BC Mental Health and Addiction Services

This has been removed by the author of this thesis for copyright reasons

5.4 Appendix D: HCR-20 (version 2) coding sheet (Webster et al., 1997) - Copyright holders Mental Health, Law, and Policy Institute, Simon Fraser University

This has been removed by the author of this thesis for copyright reasons

5.5 Appendix E: START coding sheet (Webster et al., 2009) – Copyright holders BC Mental Health and Addiction Services and St Joseph's Healthcare

This has been removed by the author of this thesis for copyright reasons

5.6 Appendix F: Data collection template for HCR-20 audit

Patient Number:

Demographic Information

Age	
Gender	
Ethnicity	
1st Diagnosis	
2nd Diagnosis	
3rd Diagnosis	
Date of admission	
Date of discharge (if applicable)	
Legal Status	

Has an HCR-20 been completed in the last 6 months? (If yes complete table opposite)

If sources of information have been identified mark which sources were used.

Ward:

HCR-20 Data

Item	Score/Completed? (Y/N)
H1	
H2	
H3	
H4	
H5	
H6	
H7	
H8	
H9	
H10	
C1	
C2	
C3	
C4	
C5	
R1	
R2	
R3	
R4	
R5	
Summary Judgments	
Risk Scenarios	
Sources of Information	
Patient Interview	
Patient Medical Records	
Staff Consultation	
Collateral Interviews	
Other	
Ward at time of assessment	
Date of assessment	

The contents of pages 86-188 have been removed to comply with UK copyright law.

These appendices contained published journal articles. The citations and dois to the published articles are listed below.

Appendix G pp. 86-101

O'Shea, L., Mitchell, A. E., Picchioni, M. M., & Dickens, G. (2013). *Moderators of the predictive efficacy of the historical, clinical and risk management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis* doi:10.1016/j.avb.2012.11.016

Appendix H pp. 102-114

O'Shea, L. E., & Dickens, G. L. (2014). Short- term assessment of risk and treatability (START): Systematic review and meta-analysis. *Psychological Assessment, 26*(3), 990-1002. doi:10.1037/a0036794

Appendix I pp. 115-124

O'Shea, L. E., Picchioni, M. M., Mason, F. L., Sugarman, P. A., & Dickens, G. L. (2014). Differential predictive validity of the historical, clinical and risk management scales (HCR–20) for inpatient aggression. *Psychiatry Research, 220*(1-2), 669-678. doi:10.1016/j.psychres.2014.07.080

Appendix J pp. 125-137

O'Shea, ,L.E., Picchioni, M. M., Mccarthy, J., Mason, F. L., & Dickens, G. L. (2015). Predictive validity of the HCR- 20 for inpatient aggression: The effect of intellectual disability on accuracy. *Journal of Intellectual Disability Research : JIDR, 59*(11), 1042. doi:10.1111/jir.12184

Appendix K pp. 138-150

O'Shea, L.,E., Picchioni, M. M., & Dickens, G. L. (2016). The predictive validity of the short- term assessment of risk and treatability (START) for multiple adverse outcomes in a secure psychiatric inpatient setting. *Assessment, 23*(2), 150-162. doi:10.1177/1073191115573301

APPENDIX L pp. 151-160

O'Shea, L. E., & Dickens, G. L. (2015). Predictive validity of the START for unauthorised leave and substance abuse in a secure mental health setting: A pseudo-prospective cohort study. *International Journal of Nursing Studies, 52*(5), 970-979. doi:10.1016/j.ijnurstu.2015.02.007

Appendix M pp. 161-175

O'Shea, L. E., & Dickens, G. L. (2015). Predictive validity of the short- term assessment of risk and treatability (START) for aggression and self-harm in a secure mental health service: Gender differences. *International Journal of Forensic Mental Health, 14*(2), 132-146. doi:10.1080/14999013.2015.1033112

Appendix N pp. 176-188

O'Shea, L. E., Picchioni, M. M., Mason, F. L., Sugarman, P. A., & Dickens, G. L. (2014). Predictive validity of the HCR- 20 for inpatient self-harm. *Comprehensive Psychiatry, 55*(8), 1937-1949. doi:10.1016/j.comppsych.2014.07.010

6 References

't Lam, K., Lancel, M., & Hildebrand, M. (2009). *Handleiding bij deShort-Term Assessment of Risk and Treatability (START): Richtlijnen bijhet beoordelen van korte termijn risico's en behandelmogelijkheden [Manual for the Short-Term Assessment of Risk and Treatability(START): Guidelines for assessment of short-term risks and treatmentopportunities (Dutch translation)]*. Assen; the Netherlands: GGZ Drenthe.

Abderhalden, C., Needham, I., Dassen, T., Halfens, R., Haug, H. J., & Fischer, J. E. (2008). Structured risk assessment and violence in acute psychiatric wards: randomised controlled trial. *British Journal of Psychiatry, 193*(1), 44-50. doi:10.1192/bjp.bp.107.045534

Abrams, R. C., Lachs, M., McAvay, G., Keohane, D. J., & Bruce, M. L. (2002). Predictors of self-neglect in community-dwelling elders. *American Journal of Psychiatry, 159*(10), 1724-1730.

Altman, D. G., & Bland, J. M. (1994a). Diagnostic tests 1: sensitivity and specificity. *British Medical Journal, 308*, 1552.

Altman, D. G., & Bland, J. M. (1994b). Diagnostic tests 2: Predictive values. *British Medical Journal, 309*(6947), 102.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders: DSM- IV* (4th ed. ed.). Washington, DC: American Psychiatric Association.

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual review of Psychology, 53*, 27-51.

Arbach-Lucioni, K., Andrés-Pueyo, A., Pomarol-Clotet, E., & Gomar-Soñes, J. (2011). Predicting violence in psychiatric inpatients: a prospective study with the HCR-20 violence risk assessment scheme. *Journal of Forensic Psychiatry and Psychology, 22*(2), 203-222. doi:10.1080/14789949.2010.530290

Arbee, F. (2014). *A retrospective record review of mental health care users who abscond from a psychiatric hospital.* (Master of Medicine in the branch of Psychiatry), University of the Witwatersrand, Johannesburg.

Augimeri, L. K. (2001). Providing effective supports. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 135-146). Burnaby, British Columbia: Mental Health, Law, and Policy Institute, Simon Fraser University.

Bain, M. R., Chalmers, J. W., & Brewster, D. H. (1997). Routinely collected data in national and regional databases--an under-used resource. *Journal of Public Health Medicine, 19*(4), 413-418.

Belfrage, H., & Douglas, K. S. (2002). Treatment effects on forensic psychiatric patients measured with the HCR-20 violence risk assessment scheme. *International Journal of Forensic Mental Health, 1*(1), 25-36.

Belfrage, H., & Fransson, G. (2001). Creating feasible plans. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 119-124). Burnaby, British Columbia: Mental Health, Law, and Policy Institute, Simon Fraser University.

The contents of pages 86-188 have been removed to comply with UK copyright law.

These appendices contained published journal articles. The citations and dois to the published articles are listed below.

Appendix G pp. 86-101

O'Shea, L., Mitchell, A. E., Picchioni, M. M., & Dickens, G. (2013). *Moderators of the predictive efficacy of the historical, clinical and risk management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis* doi:10.1016/j.avb.2012.11.016

Appendix H pp. 102-114

O'Shea, L. E., & Dickens, G. L. (2014). Short- term assessment of risk and treatability (START): Systematic review and meta-analysis. *Psychological Assessment, 26*(3), 990-1002. doi:10.1037/a0036794

Appendix I pp. 115-124

O'Shea, L. E., Picchioni, M. M., Mason, F. L., Sugarman, P. A., & Dickens, G. L. (2014). Differential predictive validity of the historical, clinical and risk management scales (HCR–20) for inpatient aggression. *Psychiatry Research, 220*(1-2), 669-678. doi:10.1016/j.psychres.2014.07.080

Appendix J pp. 125-137

O'Shea, ,L.E., Picchioni, M. M., Mccarthy, J., Mason, F. L., & Dickens, G. L. (2015). Predictive validity of the HCR- 20 for inpatient aggression: The effect of intellectual disability on accuracy. *Journal of Intellectual Disability Research : JIDR, 59*(11), 1042. doi:10.1111/jir.12184

Appendix K pp. 138-150

O'Shea, L.,E., Picchioni, M. M., & Dickens, G. L. (2016). The predictive validity of the short- term assessment of risk and treatability (START) for multiple adverse outcomes in a secure psychiatric inpatient setting. *Assessment, 23*(2), 150-162. doi:10.1177/1073191115573301

APPENDIX L pp. 151-160

O'Shea, L. E., & Dickens, G. L. (2015). Predictive validity of the START for unauthorised leave and substance abuse in a secure mental health setting: A pseudo-prospective cohort study. *International Journal of Nursing Studies, 52*(5), 970-979. doi:10.1016/j.ijnurstu.2015.02.007

Appendix M pp. 161-175

O'Shea, L. E., & Dickens, G. L. (2015). Predictive validity of the short- term assessment of risk and treatability (START) for aggression and self-harm in a secure mental health service: Gender differences. *International Journal of Forensic Mental Health, 14*(2), 132-146. doi:10.1080/14999013.2015.1033112

Appendix N pp. 176-188

O'Shea, L. E., Picchioni, M. M., Mason, F. L., Sugarman, P. A., & Dickens, G. L. (2014). Predictive validity of the HCR- 20 for inpatient self- harm. *Comprehensive Psychiatry, 55*(8), 1937-1949. doi:10.1016/j.comppsych.2014.07.010

Belfrage, H., Fransson, G., & Strand, S. (2004). Management of violent behaviour in the correctional system using qualified risk assessments. *Legal and Criminological Psychology, 9*(1), 11-22.

Berman, A. H., Bergman, H., Palmstierna, T., & Schlyter, F. (2005). Evaluation of the Drug Use Disorders Identification Test (DUDIT) in criminal justice and detoxification settings and in a Swedish population sample. *European Addiction Research, 11*(1), 22-31. doi:10.1159/000081413

Bjorkly, S., Hartvig, P., Heggen, F. A., Brauer, H., & Moger, T. A. (2009). Development of a brief screen for violence risk (V-RISK-10) in acute and general psychiatry: An introduction with emphasis on findings from a naturalistic test of interrater reliability. *European Psychiatry, 24*(6), 388-394. doi:10.1016/j.eurpsy.2009.07.004

Black, D. W., Blum, N., Pfohl, B., & Hale, N. (2004). Suicidal Behavior in Borderline Personality Disorder: Prevalence, Risk Factors, Prediction, and Prevention. *Journal of Personality Disorders, 18*(3), 226-239. doi:10.1521/pedi.18.3.226.35445

Bland, J. M., & Altman, D. G. (2000). Statistics notes. The odds ratio. *British Medical Journal, 320*(7247), 1468.

Bo, S., Abu-Akel, A., Kongerslev, M., Haahr, U. H., & Simonsen, E. (2011). Risk factors for violence among patients with schizophrenia. *Clinical psychology review, 31*(5), 711-726.

Boer, D. P., Frize, M., Pappas, R., Morrissey, C., & Lindsay, W. R. (2010). Suggested adaptations to the HCR-20 for offenders with intellectual disabilities. In L. A. Craig, W. R. Lindsay, & K. D. Browne (Eds.), *Assessment and treatment of sexual offenders with intellectual disabilities: A handbook* (pp. 177-192). Chichester: Wiley, Blackwell.

Borum, R., Swartz, M., Swanson, J., & Wiseman, S. (2001). Compliance with remediation attempts. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide*. Burnaby, BC, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., . . . de Vet, H. C. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *British Medical Journal, 326*(7379), 24-28.

Bouch, J., & Marshall, J. J. (2003). *S-RAMM Suicide Risk Assessment and Management Manual (Research Edition)*. Cardiff: The Cognitive Centre Foundation.

Bowers, L. (2014). Safewards: a new model of conflict and containment on psychiatric wards. *Journal of Psychiatric and Mental Health Nursing, 21*(6), 499-508.

Bowers, L., Banda, T., & Nijman, H. (2010). Suicide inside: a systematic review of inpatient suicides. *The Journal of nervous and mental disease, 198*(5), 315-328. doi:10.1097/NMD.0b013e3181da47e2

Bowers, L., Jarrett, M., & Clark, N. (1998). Absconding: a literature review. *Journal of Psychiatric and Mental Health Nursing, 5*, 343-353.

Bowers, L., Nijman, H., & Palmstierna, T. (2007). The attempted and actual assault scale (Attacks). *International Journal of Methods in Psychiatric Research, 16*(3), 171-176. doi:10.1002/mpr.219

Bowers, L., Stewart, D., Papadopoulos, C., Dack, C., Ross, J., Khanom, H., & Jeffery, D. (2011). Inpatient violence and aggression: a literature review: Report from the Conflict and Containment Reduction Research Programme. Retrieved from doi:http://www.kcl.ac.uk/iop/depts/hspr/research/ciemh/mhn/projects/litreview/LitRevAgg.pdf

Braithwaite, E., Charette, Y., Crocker, A. G., & Reyes, A. (2010). The predictive validity of clinical ratings of the Short-Term Assessment of Risk and Treatability (START). *International Journal of Forensic Mental Health, 9*(4), 271-281.

Burnett, J., Regev, T., Pickens, S., Prati, L. L., Aung, K., Moore, J., & Dyer, C. B. (2007). Social networks: A profile of the elderly who self-neglect. *Journal of Elder Abuse and Neglect, 18*(4), 35-49.

Caldwell, C. B., & Gottesman, I. I. (1990). Schizophrenics kill themselves too: a review of risk factors for suicide. *Schizophrenia bulletin, 16*(4), 571-589.

Carroll, A. (2009). How to make good-enough risk decisions. *Advances in psychiatric treatment, 15*(3), 192-198.

Centre for Mental Health. (2011). *Pathways to unlocking secure mental health care*. London: Centre for Mental Health.

Chesney-Lind, M. (1989). Girls' crime and woman's place: Toward a feminist model of female delinquency. *Crime and Delinquency, 35*(1), 5-29.

Choi, N. G., Kim, J., & Asseff, J. (2009). Self-Neglect and Neglect of Vulnerable Older Adults: Reexamination of Etiology. *Journal of Gerontological Social Work, 52*(2), 171-187. doi:10.1080/01634370802609239

Chu, C. M., Thomas, S. D., Ogloff, J. R. P., & Daffern, M. (2011). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) in a secure forensic hospital: Risk factors and strengths. *International Journal of Forensic Mental Health, 10*(4), 337-345.

Chu, C. M., Thomas, S. D., Ogloff, J. R. P., & Daffern, M. (2013). The short- to medium-term predictive accuracy of static and dynamic risk assessment measures in a secure forensic hospital. *Assessment, 20*(2), 230-241.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155-159.

Coid, J. W., Hickey, N., Kahtan, N., Zhang, T., & Yang, M. (2007). Patients discharged from medium secure forensic psychiatry services: reconvictions and risk factors. *British Journal of Psychiatry, 190*, 223-229. doi:10.1192/bjp.bp.105.018788

Coid, J. W., Kallis, C., Doyle, M., Shaw, J., & Ullrich, S. (2015). Identifying Causal Risk Factors for Violence among Discharged Patients. *PLoS One, 10*(11), e0142493. doi:10.1371/journal.pone.0142493

Crocker, A. G., Garcia, A., Israël, M., Hindle, Y., Gagnon, D., Venegas, C., . . . Roy, M. E. (2008). *Implementing and using a systematic risk assessment scheme to increase patient safety on a risk management unit for individuals with severe mental illness: A demonstration project*. Edmonton, Canada: Canadian Patient Safety Institute.

Csipke, E., Flach, C., McCrone, P., Rose, D., & Tilley, J. (2014). Inpatient care 50 years after the process of deinstitutionalisation. *Social Psychiatry and Psychiatric Epidemiology, 49*, 665-671.

de Lusignan, S., & van Weel, C. (2006). The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice, 23*(2), 253-263. doi:10.1093/fampra/cmi106

de Ruiter, C., & Nicholls, T. L. (2011). Protective factors in forensic mental health: A new frontier. *International Journal of Forensic Mental Health, 10*(3), 160-170. doi:10.1080/14999013.2011.600602

de Vogel, V., & de Ruiter, C. (2005). The HCR-20 in Personality Disordered Female Offenders: A Comparison with a Matched Sample of Males. *Clinical Psychology and Psychotherapy, 12*(3), 226-240.

de Vogel, V., de Vries Robbe, M., Van Kalmthout, W., & Place, C. (2012). *Female Additional Manual (FAM): Additional guidelines to the HCR-20 for assessing risk for violence in women. English version*. Utrecht; NL: Van der Hoeven Kliniek.

de Vogel, V., de Vries Robbe, M., Van Kalmthout, W., & Place, C. (2014). *Female additional manual (FAM). Additional guidelines to the HCR-20 V3 for assessing risk for violence in women. English version*. Utrecht; NL: Van der Hoeven Kliniek.

de Vries Robbe, M., de Vogel, V., & Douglas, K. S. (2013). Risk factors and protective factors: A two-sided dynamic approach to violence risk assessment. *Journal of Forensic Psychiatry and Psychology, 24*(4), 440-457.

DeClue, G., & Campbell, T. W. (2013). Calibration performance indicators for the Static-99R: 2013 Update. *Open Access Journal of Forensic Psychology, 5*, 82-88.

Department of Health. (2015). *The NHS constitution for England: the NHS belongs to us all*.

Desmarais, S. L., Collins, M., Nicholls, T., & Brink, J. (2011). *Perceptions of the Short-Term Assessment of Risk and Treatability (START) as implemented in forensic psychiatric practice* Unpublished Manuscript.

Desmarais, S. L., Nicholls, T. L., Wilson, C. M., & Brink, J. (2012). Using dynamic risk and protective factors to predict inpatient aggression: Reliability and validity of START assessments. *Psychological Assessment, 24*(3), 685-700.

DH Research and Development Directorate [England], National Institute for Social Care and Health Research [Wales], Chief Scientist Office [Scotland], & R&D Division Public Health Agency [Northern Ireland]. (2011). Governance arrangements for research ethics committees: a harmonised edition.

Dickens, G. L., & O'Shea, L. E. (2015). How short should short-term risk assessment be? Determining the optimum interval for START reassesment in a secure mental health service. *Journal of Psychiatric and Mental Health Nursing, 22*(6), 397-406.

Dickens, G. L., Picchioni, M. M., & Long, C. (2013). Aggression in specialist secure and forensic inpatient mental health care: incidence across care pathways. *Journal of Forensic Practice, 15*(3), 206-217.

Dolan, M., & Doyle, M. (2000). Violence risk prediction. *British Journal of Psychiatry, 177*(4), 303-311.

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20: Assessing risk for violence (Version 3)*. Burnaby; BC: Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., Guy, L. S., & Wilson, C. M. (2014). Historical-Clinical-Risk Management-20, Version 3 (HCR-

20V3): Development and Overview. *International Journal of Forensic Mental Health, 13*(2), 93-108.

Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment - Clinical and research applications. *Criminal Justice and Behavior, 29*(5), 617-658. doi:10.1177/009385402236735

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.), *Handbook of Violence Risk Assessment* (pp. 147-185). NY: Routledge.

Douglas, K. S., Shaffer, C., Blanchard, A. J. E., Guy, L. S., Reeves, K. A., & Weir, J. (2014). HCR-20 violence risk assessment scheme: Overview and annotated bibliography (current up to January 1, 2014). Retrieved from

Douglas, K. S., Webster, C. D., Hart, S. D., Eaves, D., & Ogloff, J. R. P. (Eds.). (2001). *HCR-20: Violence risk management companion guide*. Burnaby, BC, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University, and Department of Mental Health Law & Policy, University of South Florida.

Doyle, M., & Dolan, M. (2002). Violence risk assessment: combining actuarial and clinical information to structure clinical judgements for the formulation and management of risk. *Journal of Psychiatric and Mental Health Nursing, 9*(6), 649-657.

Dvoskin, J. A., & Heilbrun, K. (2001). Risk assessment and release decision-making: Toward resolving the great debate. *Journal of the American Academy of Psychiatry and the Law, 29*, 6-10.

Dyer, C. B., Goodwin, J. S., Pickens-Pace, S., Burnett, J., & Kelly, P. A. (2007). Self-neglect among the elderly: a model based on more than 500 patients seen by a geriatric medicine team. *American Journal of Public Health, 97*(9), 1671-1676. doi:10.2105/AJPH.2006.097113

Dyer, C. B., Kelly, P. A., Pavlik, V. N., Lee, J., Doody, R. S., Regev, T., . . . Smith, S. M. (2006). The making of a self-neglect severity scale. *Journal of Elder Abuse and Neglect, 18*(4), 13-23.

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. *British Medical Journal, 345*.

Fisher, W. H., Geller, J. L., & Pandiani, J. A. (2009). The changing role of the state psychiatric hospital. *Health Affairs, 28*(3), 676-684.

Fitzgerald, S., Gray, N. S., Alexander, R. T., Bagshaw, R., Chesterman, P., Huckle, P., . . . Snowden, R. J. (2013). Predicting Institutional Violence in Offenders with Intellectual Disabilities: The Predictive Efficacy of the VRAG and the HCR-20. *Journal of Applied Research in Intellectual Disabilities, 26*(5), 384-393.

Fliege, H., Lee, J., Grimm, A., & Klapp, B. F. (2009). Risk factors and correlates of deliberate self-harm behavior: A systematic review. *Journal of Psychosomatic Research, 66*(6), 477-493. doi:http://dx.doi.org/10.1016/j.jpsychores.2008.10.013

Flynn, G., O'Neill, C., McInerney, C., & Kennedy, H. G. (2011). The DUNDRUM-1 structured professional judgment for triage to appropriate levels of

therapeutic security: retrospective-cohort validation study. *BMC psychiatry, 11*(1), 43.

Freestone, M., Taylor, C., Milsom, S., Mikton, C., Ullrich, S., Phillips, O., & Coid, J. (2012). Assessments and admissions during the first 6 years of a UK medium secure DSPD service. *Criminal Behavior and Mental Health, 22*(2), 91-107. doi:10.1002/cbm.1823

Fujii, D. E. M., Tokioka, A. B., Lichton, A. I., & Hishinuma, E. (2005). Ethnic Differences in Prediction of Violence Risk With the HCR-20 Among Psychiatric Inpatients. *Psychiatric Services, 56*(6), 711-716.

Fulmer, T., Guadagno, L., Dyer, C. M., & Connolly, M. T. (2004). Progress in Elder Abuse Screening and Assessment Instruments. *Progress in Geriatrics, 52*(2), 297-304.

Gadon, L., Johnstone, L., & Cooke, D. (2006). Situational variables and institutional violence: A systematic review of the literature. *Clinical psychology review, 26*(5), 515-534.

Gibbons, S., Lauder, W., & Ludwick, R. (2006). Self-neglect: a proposed new NANDA diagnosis. *International Journal of Nursing Terminologies and Classifications, 17*(1), 10-18. doi:10.1111/j.1744-618X.2006.00018.x

Gottfredson, S. D., & Moriarty, L. J. (2006). Statistical Risk Assessment: Old Problems and New Applications. *Crime and Delinquency, 52*, 178-200.

Granello, D. H. (2010). The process of suicide risk assessment: Twelve core principles. *Journal of Counseling and Development, 88*(3), 363-370.

Grann, M., & Langstrom, N. (2007). Actuarial assessment of violence risk - To weigh or not to weigh? *Criminal Justice and Behavior, 34*(1), 22-36. doi:10.1177/0093854806290250

Gratz, K. L. (2001). Measurement of Deliberate Self-Harm: Preliminary Data on the Deliberate Self-Harm Inventory. *Journal of Psychopathology and Behavioral Assessment, 23*(4), 253-263.

Gray, N. S., Benson, R., Craig, R., Davies, H., Fitzgerald, S., Huckle, P., . . . Snowden, R. J. (2011). The Short-Term Assessment of Risk and Treatability (START): A prospective study of inpatient behavior. *International Journal of Forensic Mental Health, 10*(4), 305-313.

Grevatt, M., Thomas-Peter, B., & Hughes, G. (2004). Violence, mental disorder and risk assessment: Can structured clinical assessments predict the short-term risk of inpatient violence? *Journal of Forensic Psychiatry and Psychology, 15*(2), 278-292.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law, 2*(2), 293-323.

Hallett, N., Huber, J. W., & Dickens, G. L. (2014). Violence prevention in inpatient psychiatric settings: Systematic review of studies about the perceptions of care staff and patients. *Aggression and Violent Behavior, 19*(5), 502-514.

Hanson, R. K. (2008). What statistics should we use to report predictive accuracy. *Crime Scene, 15*(1), 15-17.

Harris, S. T., Oakley, C., & Picchioni, M. (2013). Quantifying violence in mental health research. *Aggression and Violent Behavior, 18*(6), 695-701.

Hart, S. D. (1998a). Psychopathy and risk for violence *Psychopathy: Theory, research and implications for society* (pp. 355-373): Springer.

Hart, S. D. (1998b). The role of psychopathy in assessing risk for violence: conceptual and methodological issue. *Legal and Criminological Psychology, 3*(1), 121-138.

Hart, S. D. (2001). Assessing and managing violence risk. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 13-26). Burnaby; BC: Mental Health, Law, and Policy Institute, Simon Fraser University.

Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: evaluating the 'margins of error' of group v. individual predictions of violence. *British Journal of Psychiatry, 49*, S60-S65. doi:10.1192/bjp.190.5.s60

Hawton, K., Sutton, L., Haw, C., Sinclair, J., & Deeks, J. J. (2005). Schizophrenia and suicide: systematic review of risk factors. *British Journal of Psychiatry, 187*(1), 9-20. doi:10.1192/bjp.187.1.9

Health & Social Care Information Centre. (2014). Inpatients formally detained in hospitals under the Mental Health Act 1983, and patients subject to supervised community treatment: Annual Report, England, 2013/2014: Health & Social Care Information Centre.

Hengeller, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents.* New York: Guilford Press.

Hodgins, S., Alderton, J., Cree, A., Aboud, A., & Mak, T. (2007). Aggressive behaviour, victimisation and crime Aggressive behaviour, victimisation and crime among severely mentally ill patients requiring hospitalisation. *British Journal of Psychiatry, 191*, 343-350.

Ireland, J. L. (2005). Exploring definitions of bullying among personality disordered patients in a maximum-secure hospital. *Aggressive Behavior, 31*, 359-373.

Ireland, J. L., & Ireland, C. A. (2003). How do offenders define bullying? A study of adult, young and juvenile offenders. *Legal and Criminological Psychology, 8*(159-173).

Iris, M., Ridings, J. W., & Conrad, K. J. (2010). The development of a conceptual model for understanding elder self-neglect. *Gerontologist, 50*(3), 303-315. doi:10.1093/geront/gnp125

James, K., Stewart, D., & Bowers, L. (2012). Self-harm and attempted suicide within inpatient psychiatric services: A review of the Literature. *International Journal of Mental Health Nursing, 21*(4), 301-309. doi:10.1111/j.1447-0349.2011.00794.x

Janes, H., Longton, G., & Pepe, M. (2009). Accommodating Covariates in ROC Analysis. *The Stata Journal, 9*(1), 17-39.

Jessor, R., Van Den Bos, J., Vanderryn, J., Costa, F. M., & Turbin, M. S. (1995). Protective factors in adolescent problem behavior: Moderator effects and developmental change. *Developmental psychology, 31*(6), 923-933.

Johnson, K. L., Desmarais, S. L., Van Dorn, R. A., & Grimm, K. J. (2015). A typology of community violence perpetration and victimization among adults with mental illness. *Journal of Interpersonal Violence, 30*(3), 522-540. doi:10.1177/0886260514535102

Kapur, N., Cooper, J., O'Connor, R. C., & Hawton, K. (2013). Non-suicidal self-injury v. attempted suicide: new diagnosis or false dichotomy? *British Journal of Psychiatry, 202*(5), 326-328. doi:10.1192/bjp.bp.112.116111

Kay, S. R., Wolkenfeld, F., & Murrill, L. M. (1988). Profiles of aggression among psychiatric patients: II. Covariates and predictors. *The Journal of nervous and mental disease, 176*(9), 547-557.

Khiroya, R., Weaver, T., & Maden, T. (2009). Use and perceived utility of structured violence risk assessments in English medium secure forensic units. *Psychiatric Bulletin, 33*(129-132).

Klonsky, E. D., & Muehlenkamp, J. J. (2007). Self-injury: A research review for the practitioner. *Journal of clinical psychology, 63*(11), 1045-1056.

Kohlman-Thompson, L. (1992). Kohlman Evaluation of Residence Skills. Bethesda, MD: American Occupational Therapy Association.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of general psychiatry, 54*(4), 337-343.

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological psychiatry, 59*(11), 990-996.

Kroppan, E., Nesset, M. B., Nonstad, K., Pedersen, T. W., Almvik, R., & Palmstierna, T. (2011). Implementation of the Short Term Assessment of Risk and Treatability (START) in a Forensic High Secure Unit. *International Journal of Forensic Mental Health, 10*(1), 7-12. doi:10.1080/14999013.2011.552368

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. doi:10.1016/S0140-6736(02)11133-0

Kung, H. C., Liu, X., & Juon, H. S. (1998). Risk factors for suicide in Caucasians and in African-Americans: a matched case-control study. *Social Psychiatry and Psychiatric Epidemiology, 33*(4), 155-161.

Kung, H. C., Pearson, J. L., & Liu, X. (2003). Risk factors for male and female suicide decedents ages 15–64 in the United States. *Social Psychiatry and Psychiatric Epidemiology, 38*(8), 419-426. doi:10.1007/s00127-003-0656-x

Lanctôt, N., & Guay, S. (2014). The aftermath of workplace violence among healthcare workers: A systematic literature review of the consequences. *Aggression and Violent Behavior, 19*(5), 492-501.

Langton, C. M., Hogue, T. E., Daffern, M., Mannion, A., & Howells, K. (2009). Prediction of institutional aggression among personality disordered forensic patients using actuarial and structured clinical risk assessment tools: Prospective evaluation of the HCR-20, VRS, Static-99, and Risk Matrix 2000. *Psychology, Crime and Law, 15*(7), 635-659.

Large, M. M., Ryan, C. J., Singh, S. P., Paton, M. B., & Nielssen, O. B. (2011). The predictive value of risk categorization in schizophrenia. *Harvard review of psychiatry, 19*(1), 25-33.

Leistico, A. R., Salekin, R. T., DeCoster, J., & Rogers, R. (2008). A Large-scale meta-analysis relating the Hare Measures of Psychopathy to antisocial conduct. *Law and Human Behaviour, 32*, 28-45.

Lerner, A. G., Bernabé-Ortiz, A., Ticse, R., Hernandez, A., Huaylinos, Y., Pinto, M. E., . . . CRONICAS Cohort Study Group. (2015). Type 2 diabetes and

cardiac autonomic neuropathy screening using dynamic pupillometry. *Diabetic Medicine, 32*(11), 1470-1478. doi:10.1111/dme.12752

Lindsay, W. R., Hogue, T. E., Taylor, J. L., Steptoe, L., Mooney, P., O'Brien, G., . . . Smith, A. H. W. (2008). Risk Assessment in Offenders with Intellectual Disability: A Comparison Across Three Levels of Security. *International Journal of Offender Therapy and Comparative Criminology, 52*(1), 90-111.

Martins, R. K., & McNeil, D. W. (2009). Review of Motivational Interviewing in promoting health behaviors. *Clinical psychology review, 29*(4), 283-293. doi:10.1016/j.cpr.2009.02.001

Mathé, E. A., Patterson, A. D., Haznadar, M., Manna, S. K., Krausz, K. W., Bowman, E. D., . . . Harris, C. C. (2014). Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. *Cancer Research, 74*(12), 3259-3270. doi:10.1158/0008-5472.CAN-14-0109

McCormack, Janus, M. D., & Burgess, A. W. (1986). Runaway youths and sexual victimization: Gender differences in an adolescent runaway population. *Child Abuse and Neglect, 10*(3), 387-395. doi:http://dx.doi.org/10.1016/0145-2134(86)90014-1

McDermott, B. E., Quanbeck, C. D., Busse, D., Yastro, K., & Scott, C. L. (2008). The accuracy of risk assessment instruments in the prediction of impulsive versus predatory aggression. *Behavioral Sciences and the Law, 26*(6), 759-777.

McGregor, K., Castle, D., & Dolan, M. (2012). Schizophrenia spectrum disorders, substance misuse, and the four-facet model of psychopathy: the relationship to violence. *Schizophrenia Research, 136*(1-3), 116-121. doi:doi: 10.1016/j.schres.2011.09.010

McKenzie, B., & Curr, H. (2005). Predicting violence in a medium secure setting: A study using the historical and clinical scales of the HCR-20. *The British Journal of Forensic Practice, 7*(3), 22-28.

McMain, S. F., & Courbasson, C. M. A. (2001). Impulse Control. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide*. Burnaby, BC, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Meichenbaum, D. (2007). Stress inoculation training: A preventative and treatment approach. In P. M. Lehrer, R. L. Woolfolk, & W. E. Sime (Eds.), *Principles and practice of stress management (Third edition)* (Third ed., pp. 497-516). NY: Guilford Press.

Mental Health Foundation. (2015). *Fundamental facts about mental health*. London, UK: Mental Health Foundation.

Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: preparing people for change* (2nd edition ed.). New York, NY: Guilford.

Monahan, J. (1992). Mental disorder and violent behavior: Perceptions and evidence. *American Psychologist, 47*(4), 511-521.

Monahan, J. (2002). The MacArthur studies of violence risk. *Criminal Behavior and Mental Health, 12*(S1), S67-S72.

Morris, J., Lora, A., McBain, R., & Saxena, S. (2012). Global Mental Health Resources and Services: A WHO Survey of 184 Countries. *Public Health Reviews, 34*(2), 1-19.

Morrissey, C., Hogue, T., Mooney, P., Allen, C., Johnston, S., Hollin, C., . . . Taylor, J. L. (2007). Predictive validity of the PCL-R in offenders with intellectual disability in a high secure hospital setting: Institutional aggression. *Journal of Forensic Psychiatry and Psychology, 18*(1), 1-15.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*(4), 783-792.

Mossman, D. (2013). Evaluating risk assessments using receiver operating characteristic analysis: rationale, advantages, insights, and limitations. *Behavioral Sciences and the Law, 31*(1), 23-39. doi:10.1002/bsl.2050

Motamed, N., Perumal, D., Zamani, F., Ashrafi, H., Haghjoo, M., Saeedian, F. S., . . . Asouri, M. (2015). Conicity Index and Waist-to-Hip Ratio Are Superior Obesity Indices in Predicting 10-Year Cardiovascular Risk Among Men and Women. *Clinical Cardiology, 38*(9), 527-534. doi:10.1002/clc.22437

Muller-Isberner, R. (2001). Changing attitudes: effecting positive and lasting changes. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 85-92). Burnaby, British Columbia: Mental Health, Law, and Policy Institute, Simon Fraser University.

National Confidential Enquiry into Suicide and Homicide by People with Mental Illness. (2014). Annual Report: England and Wales: The University of Manchester.

National Institute for Health and Clinical Excellence. (2005). *Violence: The short-term management of disturbed/violent behaviour in in-patient psychiatric settings and emergency departments*: www.nice.org.uk.

National Institute for Mental Health in England. (2004). Mental Health Policy Implementation Guide. *Retrieved 1st March 2012 from http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/Publicatio nsPolicyAndGuidance/DH_4087136*. Retrieved from

Nederlof, A. F., Muris, P., & Hovens, J. E. (2011). Threat/control-override symptoms and emotional reactions to positive symptoms as correlates of aggressive behavior in psychotic patients. *The Journal of nervous and mental disease, 199*(5), 342-347.

Nicholls, T. L., Gagnon, N., Crocker, A. G., Brink, J., Desmarais, S. L., & Webster, C. D. (2007). *START Outcomes Scale (SOS)*. Vancouver: BC Mental Health & Addiction Services.

Nicholls, T. L., Ogloff, J. R. P., & Douglas, K. S. (2004). Assessing Risk for Violence among Male and Female Civil Psychiatric Patients: The HCR-20, PCL:SV, and VSC. *Behavioral Sciences and the Law, 22*(1), 127-158.

Nicholls, T. L., Petersen, K. L., Brink, J., & Webster, C. D. (2011). A clinical and risk profile of forensic psychiatric patients: Treatment team STARTs in a Canadian service. *International Journal of Forensic Mental Health, 10*(3), 187-199.

Nijman, H. L. I., Bowers, L., Oud, N., & Jansen, G. (2005). Psychiatric nurses' experiences with inpatient aggression. *Aggressive Behavior, 31*(3), 217-227. doi:10.1002/ab.20038

Nijman, H. L. I., & Campo, J. M. L. G. À. (2002). Situational determinants of inpatient self-harm. *Suicide and Life-Threatening Behavior, 32*(2), 167-175. doi:10.1521/suli.32.2.167.24401

Nolan, K. A., Volavka, J., Czobor, P., Sheitman, B., Lindenmayer, J. P., Citrome, L. L., . . . Lieberman, J. A. (2005). Aggression and psychopathology in treatment-resistant inpatients with schizophrenia and schizoaffective disorder. *Journal of Psychiatric Research, 39*(1), 109-115. doi:http://dx.doi.org/10.1016/j.jpsychires.2004.04.010

Nonstad, K., Nesset, M. B., Kroppan, E., Pedersen, T. W., Nøttestad, J. A., Almvik, R., & Palmstierna, T. (2010). Predictive validity and other psychometric properties of the Short-Term Assessment of Risk and Treatability (START) in a Norwegian high secure hospital. *International Journal of Forensic Mental Health, 9*(4), 294-299. doi:10.1080/14999013.2010.534958

O'Brien, J. G., Thibault, J. M., Turner, L. C., & Laird-Fick, H. S. (2000). Self-neglect: an overview. *Journal of Elder Abuse and Neglect, 11*(2), 1-19.

O'Shea, L. E., & Dickens, G. L. (2014). Short-Term Assessment of Risk and Treatability (START): Systematic review and meta-analysis. *Psychological Assessment, 26*(3), 990-1002.

O'Shea, L. E., & Dickens, G. L. (2015a). Predictive validity of the Short-Term Assessment of Risk and Treatability (START) for aggression and self-harm in a secure mental health service: gender differences. *International Journal of Forensic Mental Health, 14*(2), 132-146.

O'Shea, L. E., & Dickens, G. L. (2015b). Predictive validity of the START for unauthorised leave and substance abuse in a secure mental health setting: A pseudo-prospective cohort study. *International Journal of Nursing Studies, 52*(5), 970-979.

O'Shea, L. E., & Dickens, G. L. (2015c). The HCR-20 as a measure of reliable and clinically significant change in violence risk among secure psychiatric inpatients. *Comprehensive Psychiatry, 62*, 132-140.

O'Shea, L. E., & Dickens, G. L. (2016). Role of assessment components and recent adverse outcomes in risk estimation and prediction: Use of the Short Term Assessment of Risk and Treatability (START) in an adult secure inpatient mental health service. *Psychiatry Research, 240*, 398-405.

O'Shea, L. E., Mitchell, A. E., Picchioni, M. M., & Dickens, G. L. (2013). Moderators of the predictive efficacy of the historical, clinical and risk management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis. *Aggression and Violent Behavior, 18*, 255-270.

O'Shea, L. E., Picchioni, M. M., & Dickens, G. L. (2016). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) for multiple adverse outcomes in a secure psychiatric inpatient setting. *Assessment, 23*(2), 150-162.

O'Shea, L. E., Picchioni, M. M., Mason, F. L., Sugarman, P. A., & Dickens, G. L. (2014a). Differential predictive validity of HCR-20 for inpatient aggression. *Psychiatry Research, 220*, 669-678.

O'Shea, L. E., Picchioni, M. M., Mason, F. L., Sugarman, P. A., & Dickens, G. L. (2014b). Predictive validity of the HCR-20 for inpatient self-harm. *Comprehensive Psychiatry, 55*(8), 1937-1949.

O'Shea, L. E., Picchioni, M. M., McCarthy, J. M., Mason, F. L., & Dickens, G. L. (2015). Predictive validity of the HCR-20 for inpatient aggression: The effect of intellectual disability on accuracy. *Journal of Intellectual Disability Research, 59*(11), 1042-1054.

Office for National Statistics. (2011). 2011 Census: Aggregate data (England and Wales).

Olsson, H., Strand, S., Kristiansen, L., Sjöling, M., & Asplund, K. (2013). Decreased Risk for Violence in Patients Admitted to Forensic Care, Measured With the HCR-20. *Archives of Psychiatric Nursing, 27*(4), 191-197. doi:http://dx.doi.org/10.1016/j.apnu.2013.03.004

Pepe, M. S., Janes, H., Longton, G., Leisenring, W., & Newcomb, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American journal of epidemiology, 159*(9), 882-890.

Poythress, E. L., Burnett, J., Naik, A. D., Pickens, S., & Dyer, C. B. (2006). Severe self-neglect: an epidemiological and historical perspective. *Journal of Elder Abuse and Neglect, 18*(4), 5-12.

Prentky, R. A., Janus, E., Barbaree, H., Schwartz, B. K., & Kafka, M. P. (2006). Sexually Violent Predators in the Courtroom: Science on Trial. *Psychology, Public Policy, and Law, 12*, 357-393.

Qin, P., Mortensen, P. B., Agerbo, E., Westergard-Nielsen, N., & Eriksson, T. (2000). Gender differences in risk factors for suicide in Denmark. *British Journal of Psychiatry, 177*(6), 546-550.

Rettenberger, M., Matthes, A., Boer, D. P., & Eher, R. (2010). Prospective Actuarial Risk Assessment: A Comparison of Five Risk Assessment Instruments in Different Sexual Offender Subtypes. *International Journal of Offender Therapy and Comparative Criminology, 54*(2), 169-186. doi:10.1177/0306624x08328755

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*(5), 615-620.

Rodríguez-Álvarez, M. X., Tahoces, P. G., & Cadarso-Suárez, C. (2011). Comparative study of ROC regression techniques: Applications for the computer-aided diagnostic system in breast cancer detection. *Computational Statistics and Data Analysis, 55*, 888-902.

Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research, 21*(4), 37-59.

Rutherford, M., & Duggan, S. (2007). Forensic mental health services: facts and figures on current provision. *http://www.centreformentalhealth.org.uk/pdfs/scmh_forensic_factfile_2007.pdf*. Retrieved from

Rutter, M. (1993). Resilience: Some conceptual considerations. *Journal of Adolescent Health, 14*(8), 626-631. doi:http://dx.doi.org/10.1016/1054-139X(93)90196-V

Ryan, R., Hill, S., Broclain, D., Horey, D., Oliver, S., Prictor, M., & Cochrane Consumers and Communication Review Group. (2013). Supplementary guidance for authors undertaking reviews with the Cochrane Consumers and Communication Review Group. Retrieved from http://cccrg.cochrane.org/author-resources(accessed 10/10/15) website:

Sampson, R. J., & Wilson, W. J. (2005). Toward a theory of race, crime, and urban inequality. In S. L. Gabiddon & H. T. Greene (Eds.), *Race, crime, and justice: a reader* (pp. 177-190). New York, NY: Routledge.

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test

(AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption - II. *Addiction, 88*(6), 791-804.

Sheldon, K., & Gallagher, J. (2010). S13-04 - Risk reductions in dangerous and severe personality disordered forensic in-patients measured with the HCR-20 violence risk assessment scheme. *European Psychiatry, 25*(S1), 35. doi:10.1016/s0924-9338(10)70035-3

Simera, I., Moher, D., Hoey, J., Schulz, K. F., & Altman, D. G. (2010). A catalogue of reporting guidelines for health research. *European Journal of Clinical Investigation, 40*(1), 35-53. doi:10.1111/j.1365-2362.2009.02234.x

Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: a methodological primer. *Behavioral Sciences and the Law, 31*, 8-22.

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., Yang, S., Mulvey, E. P., & RAGEE Group. (2015). Reporting guidance for violence risk assessment predictive validity studies: the RAGEE Statement. *Law and Human Behavior, 39*(1), 15-22. doi:10.1037/lhb0000090

Siris, S. G. (2001). Suicide and schizophrenia. *Journal of Psychopharmacology, 15*(2), 127-135. doi:10.1177/026988110101500209

Skeem, J., Schubert, C., Stowman, S., Beeson, S., Mulvey, E., Gardner, W., & Lidz, C. (2005). Gender and risk assessment accuracy: Underestimating women's violence potential. *Law and Human Behavior, 29*(2), 173-186.

Skegg, K. (2005). Self-harm. *Lancet, 366*, 1471-1483. doi:10.1016/S0140-6736(05)67600-3

Soloff, P. H., Lynch, K. G., & Kelly, T. M. (2002). Childhood Abuse as a Risk Factor For Suicidal Behavior in Borderline Personality Disorder. *Journal of Personality Disorders, 16*(3), 201-214. doi:10.1521/pedi.16.3.201.22542

Sourander, A., Aromaa, M., Pihlakoski, L., Haavisto, A., Rautava, P., Helenius, H., & Sillanpää, M. (2006). Early predictors of deliberate self-harm among adolescents. A prospective follow-up study from age 3 to age 15. *Journal of Affective Disorders, 93*, 87-96. doi:http://dx.doi.org/10.1016/j.jad.2006.02.015

St Andrew's Healthcare. (2016). Clinical Services Directory: Specialist care pathways leading to positive futures for all.

Steadman, H. J., Mulvey, E. P., Monahan, J., Robbins, P. C., Appelbaum, P. S., Grisso, T., . . . Silver, E. (1998). Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods. *Archives of general psychiatry, 55*(5), 393-401.

Steinert, T. (2002). Prediction of inpatient violence. *Acta Psychiatrica Scandinavica, 106*, 133-141. doi:10.1034/j.1600-0447.106.s412.29.x

Strub, D. S., Douglas, K. S., & Nicholls, T. L. (2014). The Validity of Version 3 of the HCR-20 Violence Risk Assessment Scheme Amongst Offenders and

Civil Psychiatric Patients. *International Journal of Forensic Mental Health, 13*(2), 148-159.

Sugarman, P. A., & Dickens, G. L. (2015). The evolution of secure and forensic mental healthcare. In G. L. Dickens, P. A. Sugarman, & M. M. Picchioni (Eds.), *Handbook of Secure Care* (pp. 1-14). London, UK: Royal College of Psychiatrists.

Swanson, J. W., Van Dorn, R. A., Monahan, J., & Swartz, M. S. (2006). Violence and leveraged community treatment for persons with mental disorders. *American Journal of Psychiatry, 163*(8), 1404-1411. doi:10.1176/ajp.2006.163.8.1404

Szmukler, G., & Rose, N. (2013). Risk assessment in mental health care: Values and costs. *Behavioral Sciences and the Law, 31*(1), 125-140.

Tengström, A., Hodgins, S., Müller-Isberner, R., Jöckel, D., Freese, R., Özokyay, K., & Sommer, J. (2006). Predicting Violent and Antisocial Behavior in Hospital Using the HCR-20: The Effect of Diagnoses on Predictive Accuracy. *International Journal of Forensic Mental Health, 5*(1), 39-53. doi:10.1080/14999013.2006.10471229

Torrey, E. F., Kennard, A. D., Eslinger, D., Lamb, R., & Pavle, J. (2010). *More mentally ill persons are in jails and prisons than hospitals: a survey of the states*. Arlington, VA: Treatment Advocacy Center.

Troquete, N., van den Brink, R., Beintema, H., Mulder, T., van Os, T., Schoevers, R., & Wiersma, D. (2013). Risk assessment and shared care planning in out-patient forensic psychiatry: Cluster randomised controlled trial. *British Journal of Psychiatry, 202*(5), 365-371.

Tsigebrhan, R., Shibre, T., Medhin, G., Fekadu, A., & Hanlon, C. (2014). Violence and violent victimization in people with severe mental illness in a rural low-income country setting: a comparative cross-sectional community study. *Schizophrenia Research, 152*(1), 275-282. doi:10.1016/j.schres.2013.10.032

Tyrer, P., Cooper, S., Herbert, E., Duggan, C., Crawford, M., Joyce, E., . . . Maden, T. (2007). The Quantification of Violence Scale: a simple method of recording significant violence. *International Journal of Social Psychiatry, 53*(6), 485-497.

Verbrugge, H. M., Goodman-Delahunty, J., & Frize, M. C. J. (2011). Risk Assessment in Intellectually Disabled Offenders: Validation of the Suggested ID Supplement to the HCR-20. *International Journal of Forensic Mental Health, 10*(2), 83-91.

Walsh, B. H., Broadhurst, D. I., Mandal, R., Wishart, D. S., Boylan, G. B., Kenny, L. C., & Murray, D. M. (2012). The metabolomic profile of umbilical cord blood in neonatal hypoxic ischaemic encephalopathy. *PLoS One, 7*(12), e50520. doi:10.1371/journal.pone.0050520

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence (version 2)*. Burnaby, British Columbia: Mental Health, Law and Policy Institute, Simon Fraser University.

Webster, C. D., Eaves, D., Douglas, K. S., & Wintrup, A. (1995). *The HCR-20 scheme: The assessment of dangerousness and risk*. Burnaby, British Columbia: Simon Fraser University and British Columbia Forensic Psychiatric Services Commission.

Webster, C. D., Eaves, D., & Halpin, P. (2001). Building stable environments. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff

(Eds.), *HCR-20 violence risk management companion guide* (pp. 125-134). Burnaby, British Columbia: Mental Health, Law, and Policy Institute, Simon Fraser University.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Manual for the Short-Term Assessment of Risk and Treatability (START) (Version 1.1)*. Coquitlam, BC: British Columbia Mental Health & Addiction Services.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Middleton, C. (2004). *Short-term assessment of risk and treatability (START)*. Port Coquitlam, BC: St. Joseph's Healthcare Hamilton, Ontario, and Forensic Psychiatric Services Commission.

Webster, C. D., Nicholls, T. L., Martin, M. L., Desmarais, S. L., & Brink, J. (2006). Short-Term Assessment of Risk and Treatability (START): The case for a new structured professional judgment scheme. *Behavioral Sciences and the Law, 24*(6), 747-766.

Whittington, R., Bjørngaard, J. H., Brown, A., Nathan, R., Noblett, S., & Quinn, B. (2014). Dynamic relationship between multiple START assessments and violent incidents over time: a prospective cohort study. *BMC psychiatry, 14*, 323. doi:10.1186/s12888-014-0323-7

Williams, D. R., & Williams-Morris, R. (2000). Racism and mental health: the African American experience. *Ethnicity and Health, 5*, 243-268. doi:10.1080/713667453

Wilson, C. M., Desmarais, S. L., Nicholls, T. L., & Brink, J. (2010). The role of client strengths in assessments of violence risk using the Short-Term Assessment of Risk and Treatability (START). *International Journal of Forensic Mental Health, 9*(4), 282-293.

Wilson, C. M., Desmarais, S. L., Nicholls, T. L., Hart, S. D., & Brink, J. (2013). Predictive validity of dynamic factors: assessing violence risk in forensic psychiatric inpatients. *Law and Human Behavior, 37*(6), 377-388.

Winship, G. (2011). Absconding and locking ward doors: evidence from the literature. *Journal of Psychiatric and Mental Health Nursing, 18*, 89-93.

Witt, K., van Dorn, R., & Fazel, S. (2013). Risk factors for violence in psychosis: systematic review and meta-regression analysis of 110 studies. *PLoS One, 8*(2), e55942. doi:10.1371/journal.pone.0055942

World Health Organisation. (1992). *The ICD-10 classification of mental and behavioural disorders*. Geneva: WHO.

World Health Organisation. (2011). Mental Health Atlas 2011. Geneva, Switzerland: WHO Press.

World Health Organisation. (2014). Mental Health Atlas 2014. Geneva, Switzerland: WHO Press.

World Health Organization. (2014). Preventing suicide: a global imperative. Geneva, Switzerland: WHO.

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. *Psychological bulletin, 136*(5), 740-767.

Yudofsky, S. C., Silver, J. M., Jackson, W., Endicott, J., & Williams, D. (1986). The Overt Aggression Scale for the objective rating of verbal and physical aggression. *American Journal of Psychiatry, 143*(1), 35-39.

Zimnavoda, T., Weinblatt, N., & Katz, N. (2002). Validity of the Kohlman
Evaluation of Living Skills (KELS) with Israeli elderly individuals living in
the community. *Occupational Therapy International, 9*(4), 312-325.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC)
plots: a fundamental evaluation tool in clinical medicine. *Clinical
Chemistry, 39*(4), 561-577.